ISOLATION AND CHARACTERIZATION OF THE *csa* OPERON (ETEC-CS4 PILI) AND METHODS OF USING SAME

Related Applications

This application claims priority to U.S. Provisional Patent Application No. 60/198,626, filed April 20, 2000, which is hereby incorporated by reference in its entirety.

Background of the Invention

Field of the Invention

The disclosed invention relates to the isolation and characterization of the *csa* operon, which encodes the CS4 pili and its use as an immunogenic agent with utility in preventing ETEC colonization of a subject.

Description of the Related Art

Human ETEC strains are a major cause of diarrhea in infants and young children in developing countries (Black et al., *Lancet*, I: 141-143 (1981); Levine, *J. Infect. Dis.*, 155:377-389 (1987); Qadri, et al., *J. Clin. Microbiol.*, 38:27-31 (2000)), which account for a high rate of infantile morbidity and mortality. Human ETEC strains are also a major cause of travelers' diarrhea. (Black, *Rev. Infect. Dis.*, 8S:S131-S135 (1986); DuPont et al., *N. Engl. J. Med.*, 285:1520-1521 (1976); Hyams et al., *N. Engl. J. Med.*, 325:1423-1428 (1991); Merson, et al., *N. Engl. J. Med.*, 294:1299-1305 (1976)). ETEC infection is characterized by watery diarrhea often accompanied by low-grade fever, abdominal cramps, malaise and vomiting.

ETEC strains colonize the small bowel lumen by means of surface pili called colonization factor antigens (CFA), and coli surface antigens (CS), and cause diarrhea through the action of heat labile (LT) and/or heat stable (ST) enterotoxins. ETEC fimbriae are proteinaceous filaments exhibiting different morphologies such as rigid rod like shapes of 2-7 nm in diameter, fibrilar thin flexible wiry structures, or bundles. (Gaastra et al., *Trends. Microbiol.*, 4:444-452 (1996)).

Human ETEC strains display a variety of over 20 serologically distinct pili on their cell surfaces. The most common human ETEC strains express CFA/I, CFA/II and CFA/IV (Levine, et al., "Fimbrial vaccines," *In* P. Klemm (ed.), Fimbriae: adhesion, biogenics, genetics and vaccines, Boca Raton: CRC Press, 1994; McConnell, et al., *Epidemiol. Infect.*, 106: 477-484 (1991)). CFA/I produces a single type of fimbriae, while CFA/II and CFA/IV strains produce several types of coli surface antigens. CFA/II strains express CS1, CS2 and CS3; and CFA/IV strains (originally called PCF8775) express the nonpilus antigen CS6 either alone or together with CS4 or CS5 fimbria. (McConnell, et al., *Infect. Immun.*, 56:1974-1980 (1988); McConnell, et al., *FEMS Microbiol. Lett.*, 52:105-108 (1989); Svennerholm, et al., *Infect. Immun.*, 56:523-528 (1988); Thomas, et al., *J. Gen. Microbiol.*, 131:2319-2326 (1985)). The occurrence of $CS4^+CS6^+$ producing strains is restricted to serotype O25:H42. (McConnell, et al., *Infect. Immun.*, 56:1974-1980 (1988); Willshaw, et al., *FEMS Microbiol. Lett.*, 49:473-478 (1988); Willshaw, et al., *FEMS Microbiol. Lett.*, 56:255-260 (1990); Willshaw, et al., *FEMS Microbiol. Lett.*, 66:125-129 (1991)).

The CS4 pili is rigid, 7 nm in diameter, and is composed of subunits with a molecular mass of 17.0 kDa. (Knutton, et al., *Infect. Immun.*, 57:3364-3371 (1989); McConnell, et al., *Infect. Immun.*, 56:1974-1980 (1988); Wolf, et al., *Infect. Immun.*, 57:164-173 (1989)). Because of their epidemiological importance and due to the fact that cross protection does not occur between strains of ETEC expressing different fimbriae, at least these CFA/I and CS1-CS6 fimbrial types must be included in a broad spectrum ETEC vaccine, (Gaastra et al., *Trends. Microbiol.*, 4:444-452 (1996); Levine, *J. Pediatr. Gastroenterol. Nutr.*, In Press (2000); Levine, et al., "Fimbrial vaccines," *In* P. Klemm (ed.), Fimbriae: adhesion, biogenics, genetics and vaccines, Boca Raton: CRC Press, 1994). Of these seven important fimbriae, only the genes encoding CS4 have not been cloned and sequenced.

The genes that are required for the expression of functional pili are characteristically linked in gene clusters (Sakellaris and Scott, *Mol. Microbiol.*, 30:681-687 (1998)), and consist of the structural genes, assembly cassette genes and regulatory genes. The assembly cassette genes include chaperone and usher genes. The chaperone protein is thought to bind to fimbrial subunit proteins in the periplasmic space and

prevent premature folding and degradation. The usher proteins are outer membrane proteins that serve as pores for the transport and assembly of the fimbriae. The structural gene encodes for the pilin protein that forms the fimbriae that is composed of repeated subunits of the pilin protein. Some fimbriae such as, CFA/I, CS1 and CS2, contain a minor pilin protein which is associated with the pili tip, that is probably involved in the attachment of the bacteria to the cell receptors. (Sakellaris, et al., *Proc. Natl. Acad. Sci. U.S.A.*, 96:12828-12832 (1999)). Fimbria expression is controlled by genes such as *rns* and *cfaD* that are similar to the *araC* family of transcriptional regulators that positively regulate transcription. (Grewal et al., *Vaccine*, 11:221-226 (1993); de Haan et al., *FEMS Microbiol. Lett.*, 67:341-346 (1991); 19, 21, Savelkoul, et al., *Microb. Pathog.*, 8:91-99 (1990)).

Genes encoding ETEC fimbriae have been located on large plasmids [CFA/I, (Hamers et al., *Microb. Pathog.*, 6:297-309 (1989); 20); CS1, (Froehlich et al., *Mol. Microbiol.*, 12:387-401 (1994)); CS3, (Jalajakumari, et al., *Mol. Microbiol.*, 3:1685-1695 (1989), Manning, et al., *Mol. Gen. Genet.*, 200:322-327 (1985)); CS5, (Duthy, et al., *J. Bacteriol.*, 181:5847-5851 (1999)) and CS6, (Wolf, et al., *FEMS Microbiol. Lett.*, 148:35-42 (1997))], or on the chromosome [CS2, (Froehlich et al., *Mol. Microbiol.*, 12:387-401 (1994), Froehlich et al., *Infect. Immun.*, 63:4849-4856 (1995))]. Early experiments to locate the CS4 encoding genes revealed disparate results without conclusive localization, (Sommerfelt, et al., *Microb. Pathog.*, 11:297-304 (1991); Sommerfelt, et al., *Infect. Immun.*, 60:3799-3806 (1992); Sommerfelt, et al., *J. Clin. Microbiol.*, 30:1823-1828 (1992); Willshaw, et al., *FEMS Microbiol. Lett.*, 56:255-260 (1990); Wolf, et al., *Infect. Immun.*, 57:164-173 (1989)).

Because the attachment of ETEC strains to intestinal cells is crucial for establishment of infection, the prevention of disease is based mainly upon immune responses against the pili that interfere with the attachment process. Studies performed with 14 healthy human volunteers who ingested $5x10^8$ *E. coli* E24377A (O139:H28), a CS1 and CS3 producing strain, showed that all 14 became colonized, 9 developed a typical diarrheal syndrome, and 6 of these ill persons manifested a significant increase in serum IgG antibody to purified CS1 and CS3 antigens. Levine, M.M., et al., *Infect.*

Immun., 44:409-420 (1984). These results suggest that the fimbria play a role in pathogenesis, and stimulate an immune response.

Passive protection against ETEC infection was demonstrated in a clinical trial done at the University of Maryland, which demonstrated that oral prophylaxis with hyperimmune anti CFA/I immunoglobulin provided 90% protection against diarrhea caused by oral challenge with 10^9 cfu of ETEC strain H10407, a CFA/I producing strain. Tacket, et al., N. Eng. J. Med., 318: 1240 (1988). In another study oral immunization with purified CS1 and CS3 antigens encapsulated in biodegradable polymer microspheres were used to induce the development of IgA anti CS ASC (Antibody Secreting Cells) and jejunal fluid secretory IgA anti CS in 50% of the vaccinees. Levine, et al., Fimbriae (pili) adhesions as vaccines, in *Protein-Carbohydrate Interactions in Biological Systems. The molecular Biology of Microbial Pathogenicity*, Lark, et al., Eds., Academic Press, London, p 154, 1986. This exposure also protected 30% of vaccinees from diarrhea following challenge with the virulent ETEC E24377A ($CS1^+CS3^+LT^+ST^+$) strain. A protective efficacy of 75% was demonstrated by immunization with the attenuated ETEC strain. Feeding volunteers with $5x10^{10}$ live *E. coli* E1392-75-2A (O6:H16) a $CS1^+$ $CS3^+$ LT^- ST^- strain, induced significant rise in intestinal fluid secretory IgA antibodies to CS1 and CS3 fimbria, and conferred protection to 9/12 volunteers that were challenged with a virulent heterologous serotype strain ETEC E24377A (O139:H28).

Another approach to develop an ETEC vaccine is to immunize with killed mixed ETEC strains. This type of vaccine is based on the fact that prior infection with an ETEC strain elicits protective immunity against a clinical illness that might be caused from subsequent exposure to the homologous strain. Oral immunization of children and adult volunteers with such a vaccine resulted in significant intestinal IgA responses against the CFA and CS components of the various strains. The vaccine induced high level of intestinal IgA antibody, IgA antibody ASC in the blood, and serum antibodies towards the colonization factor antigens. Jertborn, et al., Vaccine 16:255 (1997). Those results indicate that antibodies towards ETEC pilis provide protection against diarrhea caused by ETEC strains.

A combined vaccine against diarrheal disease caused by *Shigella, Salmonella* and ETEC has been proposed. These vaccines are composed of live attenuated *Shigella* or *Salmonella* strains that express ETEC fimbriae. The recombinant bacteria colonize the intestine and induced the mucosal and systemic immune response against the bacteria and the pili. Noriega et al., (*Infect. Immun.*, 64:23-27 (1996)) co-expressed CFA/I and CS3 in *Shigella* CVD1203, immunized guinea pigs and mice, and showed that the immunized animals developed high titer of tears secreted IgA (similar to mucosal sIgA), and serum IgG antibodies, towards the *Shigella* LPS and the fimbrial antigens. The ETEC human recombinant LT (K63) gene was expressed in *Shigella* CVD1204, a guanine dependent strain, and showed production of sIgA and serum IgG toward the *Shigella* LPS as well as towards the LT-A and LT-B subunits of the LT enterotoxin, following immunization of guinea pigs. Koprowski, et al., (*Infect. Immun.*, 68:4884-92 (2000)) co-expressed in CVD1204 the CFA/I and LT (K63) antigens, and demonstrates production of antibodies of the sIgA and serum IgG toward the *Shigella* LPS, the CFA/I pili antigens and the LT antigens. Altboum et al, (Attenuated *Shigella flexneri* 2a Δ*guaBA* strain CVD 1204 expressing ETEC CS2 and CS3 fimbriae as a live mucosal vaccine against *Shigella* and enterotoxigenic *Escherichia coli* infection, In press) immunized guinea pigs with a mixture of CVD1204 strains expressing ETEC CS2 and CS3 fimbria. All the immunized animals developed tears sIgA and serum IgG antibodies against *Shigella* LPS, CS2 and CS3 antigens, agglutinating antibodies against *Shigella* and ETEC CS2 and CS3 strains, and were protected against keratoconjunctivitis caused by eye challenge with the virulent *S. flexneri 2a* 2457T strain. Those results indicate that a combined immunization with live attenuated *Shigella* strains expressing ETEC fimbria might induce protection against Shigellosis and ETEC infection.

Summary of the Invention

The disclosed invention relates to compositions and methods of using *csa* operon products and the nucleotide and amino acid sequences encoded thereby. One embodiment relates to an immunogenic composition comprising a recombinant product of a *csa* operon and a carrier. Various aspects of this embodiment relate to

compositions in which the recombinant product of the *csa* operon is CsaA, CsaB, CsaC, CsaD, CsaE, or a product that is at least 95% homologous to anyone of these *csa* operon products. Additionally, the recombinant product of the *csa* operon can comprise the *csa* operon itself.

Another embodiment relates to an isolated nucleotide sequence comprising a *csa* operon or a functional fragment thereof. Another embodiments relates to a purified polypeptide sequence expressed from a recombinant *csa* operon.

Also disclosed are various methods of using the *csa* operon, products, and fragments thereof. One embodiment teaches a method of generating an immune response, comprising providing an immunogenic composition to a subject, wherein the immunogenic composition comprises a *csa* operon, a functional fragment thereof, or product thereof, and contacting the subject with the immunogenic composition, whereby an immune response is generated in the subject.

Another embodiment relates to a method of producing a polypeptide product from a *csa* operon or functional fragment thereof, comprising providing the *csa* operon in an expression vector, introducing the expression vector into a host cell, such that a recombinant host cell is produced, and subjecting the recombinant host cell to conditions such a protein from the *csa* operon is expressed.

Additional embodiment encompass cells containing recombinant the *csa* operon or fragments thereof and vectors comprising the *csa* operon or fragments thereof.

Brief Description of the Drawings

FIGURE 1 . Plasmid constructs containing the entire *csa* operon. Plasmid pKS-CSA-I contains the *csa*B, *csa*C, *csa*E, *csa*D' genes, and IS1 element, (1A). Plasmid pKS-CSA-II contains the *csa*A genes and upstream sequences flanked with IS21 element.(1B). Schematic representation of the *csa* operon, (1C).

FIGURE 2. The CS4 fimbriae expression plasmid. Schematic description of the stabilized plasmid pGA2, (Fig.2A), that is maintained with 15 copies per cell, and contained the the *hok-sok* post-segregational killing system, the *parA* and *parM* plasmid partitioning system; the *aph* allele for resistance to kanamycin, and multiple cloning

sites. Fig. 2B, Plasmid pGA2-CS4 that contained the cloned CS4 fimbriae encoding genes, *csa*A , *csa*B, *csa*C, and *csaE.*

FIGURE 3. Alignment of the amino acid sequence of CsaE to other homologous proteins. The predicted amino acid sequence of CsaE (SEQ ID NO:28) based on the gene sequence was used in a BLAST search for other homologous proteins. CfaE (SEQ ID NO:29) is the tip adhesion moiety of CS2; CooD (SEQ ID NO:30) is the tip adhesion moiety of CS1; and TsaD (SEQ ID NO:31) is the adhesion protein from *Salmonella typhi.* CotD (SEQ ID NO:32) is the spore coat protein of *Bacillus subtilis.*

FIGURE 4. Alignment of the amino acid sequence of CsaB to other ETEC fimbrial subunit proteins. The predicted amino acid sequence of CsaB (SEQ ID NO:33) based on the DNA sequence was used in a BLAST search to identify homologous proteins. The other fimbrial subunits were CfaA (SEQ ID NO:34), CooA (SEQ ID NO:35), CotA (SEQ ID NO:36), CsuA1 (SEQ ID NO:37), CsuA2 (SEQ ID NO:38), CsdA (SEQ ID NO:39), and CsbA (SEQ ID NO:40).

Detailed Description of the Preferred Embodiments

The description below relates to the *csa* operon, which encodes the proteins required for the production of CS4 antigen, and the generation of immunogenic compositions containing products or fragments of the *csa* operon. Methods of making and using the disclosed immunogenic compositions are also described.

To generate the disclosed immunogenic compositions, the *csa* operon has been sequenced and cloned into an expression vector. The genes of the *csa* operon have been expressed in *Shigella* CVD1204, resulting in the production of CS4 in this recombinant organism. Preliminary results from immunization of guinea pigs with this recombinant strain indicate the development of an immune response against both *Shigella* and the CS4 pili. Accordingly, the work described herein relating to the *csa* operon is applicable to the formulation and administration of immunogenic compositions with activity against a variety of diseases.

Organization of the *csa* Operon

The *csa* operon is located on a DNA fragment of approximately ten (10) kilobases in length. The operon comprises 5 genes that are flanked on both sides by insertion elements. This structure is similar to the pathogenicity island described in FIGURE 1.

From the entire fragment, 7245 base pairs were sequenced on both strands and the data are presented in FIGURE 2. The size of the *csa* operon was found to be approximately 6099 base pairs in length and encompasses five open reading frames.

Sequence homology to genes encoding the subunits of other ETEC fimbriae for which biochemical analysis is available indicates the following functions for each of the open reading frames. The *csaA* gene encodes the CsaA protein, which is hypothesized to be a periplasmic chaperon-like protein. The *csaB* gene encodes the CsaB protein, which is hypothesized to be the major pilin subunit. Amino terminal sequence analysis of the CS4 pilin subunit confirms the role for this gene. The *csaC* gene encodes the CsaC protein is hypothesized to be a membrane usher protein. The *csaD* gene encodes a truncated CsaD, which is hypothesized to be a regulatory protein. The *csaE* gene encodes the CsaE protein, which is hypothesized to be a tip-associated pilin protein. The amino acid sequences of the CsaA-E proteins are described in FIGURE 3.

The location of the individual genes in the *csa* operon, and the properties of the CsaA-E proteins, are described in Table 1.

TABLE 1
Properties of the *csa* operon.

Gene	Location (bp)	PRO	properties of the CS4 proteins:			
			No. of AA	Calculated MW	Theoretical pI	Signal peptide, No. of AA
csaA	283-999	CsaA	238	27305.6	9.29	19
csaB	1028-1531	CsaB	167	17343.9	6.56	23
csaC	1589-4192	CsaC	867	97686.93	8.42	22
csaE	4196-5281	CsaE	361	40102.4	8.75	23
csaD	5790-6119	CsaD	109	Missing the N-terminal 48 amino acids.		

The amino acid sequences of the CsaA-E proteins share homology to similar proteins from other ETEC fimbriae. A comparison of protein homologies is shown in Table 2.

TABLE 2

Homology of CsaA-E proteins to other ETEC fimbriae proteins.

CS4 Pro	Homology to:					
	Pili	Protein	NCBI no.	Identities	Positive	Gaps
CsaA	CFA/I	CfaA	M55661	208/238 (87%)	218/238 (91%)	
	CS1	CooB	P25731	124/221 (56%)	161/221 (72%)	4/221 (1%)
	CS2	CotB	S57934	111/238 (46%)	157/238 (65%)	3/238 (1%)
CsaB	CS4	CsfA	X97493	109/134 (81%)	110/134 (81%)	
	CFA/I	CfaB	P02971	99/170 (58%)	122/170 (71%)	3/170 (1%)
	CS14	CsuA1	X97491	95/164 (57%)	116/164 (69%)	
	CS1	CooA	P25730	85/168 (50%)	108/168 (63%)	4/168 (2%)
	CS2	CotA	S57935	76/167 (45%)	105/167 (62%)	3/167 (1%)
	CS14	CsuA2	X97492	67/132 (50%)	84/132 (62%)	1/132 (1%)
	CS19	CsdA	X97494	56/131 (42%)	75/131 (56%)	1/132 (0%)
	CS17	CsbA	X97495	54/132 (40%)	75/132 (55%)	1/132 (0%)
CsaC	CFA/I	CfaC	P25733	800/868 (92%)	821/868 (94%)	
	CS1	CooC	S49538	531/841 (63%)	659/841 (78%)	3/841 (0%)
	CS2	CotC	S57936	469/839 (55%)	610/839 (71%)	3/839 (0%)
CsaE	CFA/I	CfaE	P25734	268/361 (74%)	293/361 (80%)	1/361 (0%)
	CS1	CooD	S49539	177/329 (59%)	224/329 (67%)	11/329 3%)
	CS2	CotD	S57937	161/339 (47%)	218/339 (63%)	7/339 (2%)
CsaD	CFA/I	CfaD	P25393	90/101 (89%)	98/101 (96%)	
		RNS	P16114	89/101 (88%)	98/101 (96%)	
		CSVR	P3460	75/103 (72%)	88/103 (84%)	
	AAF/I	AGGR	P43464	57/101 (56%)		

The *csa* operon encodes the synthesis of ETEC CS4 fimbriae. This operon was cloned from ETEC E11881A, a CS4 producer strain. A nucleic acid fragment comprising 7239 base pairs was sequenced (in both directions). The results indicate that the *csa* operon is located on a nucleic acid fragment 6095 base pairs in length. (Accession No. AF296132).

The sequence of the *csa* operon was analyzed. The analysis of the operon indicated that the *csa* operon encodes five proteins. These proteins are the fimbriae

structural protein (CsaB), the tip associated protein (CsaE), a chaperon-like protein (CsaA), a usher-like protein (CsaC), and a truncated regulatory protein (CsaD).

The CsaB protein consist of 167 amino acids, (23 of which comprise a signal peptide), producing an ~17 kDa peptide. The amino acid sequence of the CsaB protein shares homology with other ETEC fimbriae proteins. For example, CsaB is 71% homologous to the CfaB protein of CFA/I, 69% and 62% homologous to the two CS14 structural proteins, 63% homologous to the CS1 structural protein, 62% homologous to the CS2 structural protein, 56% homologous to the CS19 structural protein, and 55% homologous to the CS17 structural protein.

The CsaE protein is believed to be a tip associated protein based on its homology to known tip proteins of other ETEC fimbriae. It is a protein of 361 amino acids, 23 amino acids of which are cleaved to produce a globular ~40 kDa protein. The amino acid sequence of the CsaE protein shares homology to CFA/I, CS1 and CS2 pili tip proteins, of 80%, 67% and 63%, respectively. The fimbrial assembly proteins CsaA and CsaC share homology to similar proteins from CFA/I, CS1 and CS2 pili.

Csa is believed to be a regulatory protein based on its homology to other known ETEC regulatory proteins. Its 48 terminal amino acids were deleted as a result of an insertion of an IS1 element. In addition, two frame shift mutations following 100 amino acids resulted in a stop codon. ETEC CFA/I strain E7473 contains a truncated CfaD' like protein (144 amino acids out of 265), which also contains a stop codon and various frame shift mutations. (Jordi, et al., *DNA Seq.*, 2:257-263 (1992)). The position of the frame shift mutation in *cfa*D' (Accession AAC41418) is at the same region as in *csaD'*.

Upstream to the *csaA* gene (approximately 3.5 kb) there is an additional IS21 element, rendering the *csa* operon flanked between the two insertion elements, which is characteristic of a mobile structure. The size of this pathogenicity island-like structure is about 10,500 base pairs and its G+C ratio is lower than that of *E. coli* (38.8% versus 50.8%).

Other ETEC fimbrial operons are carried on similar mobile structures. The CS1 operon contains IS sequences on both sides (IS150 (Accession: X62495), and IS2 (Accession No.: X76908)(Froehlich, et al., *Mol. Microbiol.*, 12:387-401 (1994)). The CS2 operon contains at its upstream site IS3 and IS1 DNA sequences, (Accession No.:

Z47800)(Froehlich, et al., *Infect. Immun.*, 63:4849-4856 (1995)). The CS6 pili contains IS elements, upstream IS91 and IS102, and downstream IS629 and IS3, (Accession U04844)(Wolf, et al., *FEMS Microbiol. Lett.*, 148:35-42 (1997)). The CS5 pili operon contains IS1 sequences at the upstream position that are 99% homologous to the IS1 sequences of the *csa* operon, and at its downstream site, IS30 sequences. (Accession AJ224079)(Duthy, et al., *J. Bacteriol.*, 181:5847-5851 (1999)).

The CS4 major fimbrial protein shares a high degree of amino acid sequence homology to CFA/I, CS1 and CS2 fimbrial proteins. Cross-reaction between antibodies against CS4 and CFA/I, CS1, CS2, and CS17 fimbriae were described by McConnell et al 1989. (McConnell, et al., *FEMS Microbiol. Lett.*, 52:105-108 (1989)).

The high homology between the structural proteins of CS4 and CFA/I fimbriae resulted in antibodies that cross-reacted with both fimbriae. Monoclonal antibodies against CFA/I cross reacted with CS4 and inhibited the binding of both fimbriae containing strains to human jejunal enterocytes and to Caco-2 cells. These antibodies also inhibited hemagglutination and conferred passive protections against fluid accumulation in rabbit ileal loops caused by infection of both ETEC strains. (Rudin, et al., *Microb. Pathog.*, 21:35-45 (1996)). Moreover, it has been hypothesized that immunization with purified CFA/I and CS4 fimbriae may prime and boost immune response against the homologous and heterologous fimbriae. (Rudin and Svennerholm, *Microb. Pathog.*, 16:131-139 (1994); Rudin, et al., *Epidemiol. Infect.*, 119:391-393 (1997)).

Nucleotide Sequences relating to the *csa* operon

Having identified the *csa* operon and the genes encoded thereby, this knowledge has been used to produce useful immunogenic compositions. As is discussed more fully in the Examples below, the genes of the *csa* operon were cloned, sequenced and expressed. Polynucleotide molecules encoding the proteins of the *csa* operon and their sequences are provided below.

Representative polynucleotide molecules encoding the proteins of the *csa* operon (SEQ ID. NO.: 27) include sequences comprising *csaA* (SEQ. ID. NO.: 1), *csaB* (SEQ. ID. NO.: 3), *csaC* (SEQ. ID. NO.: 5), *csaD* (SEQ. ID. NO.: 7), and *csaE* (SEQ.

ID. NO.: 9). Polynucleotide molecules encoding the proteins of the *csa* operon include those sequences resulting in minor genetic polymorphisms, differences between strains, and those that contain amino acid substitutions, additions, and/or deletions.

In some instances, one can employ such changes in the sequence of a recombinant *csa* operon-encoded protein to substantially decrease or increase the biological activity of a particular *csa* operon-encoded protein relative to the activity of the corresponding wild-type *csa* operon-encoded protein. Such changes can also be directed towards an endogenous *csa* operon encoded sequence using, for example, various molecular biological techniques to alter the endogenous gene and therefore its protein product.

Nucleotide sequences encoding *csa* operon proteins can be used to identify polynucleotide molecules encoding other proteins with biological functions similar to that of the *csa* operon. Complementary DNA molecules encoding *csa* operon-like proteins can be obtained by constructing a cDNA library from mRNA from eukaryotic cells or a DNA library from other prokaryotic organisms. DNA molecules encoding *csa* operon-like proteins can be isolated from such a library using the sequences disclosed herein with standard hybridization techniques or by the amplification of sequences using polymerase chain reaction (PCR) amplification.

In a similar manner, genomic DNA encoding *csa* operon protein homologs can be obtained using probes designed from the sequences disclosed herein. Suitable probes for use in identifying *csa* operon produced protein homologue sequences can be obtained from *csa* operon-specific sequences. Alternatively, oligonucleotides containing specific DNA sequences from a *csa* operon-coding region can be used to identify related *csa* clones. One of ordinary skill in the art will appreciate that the regulatory regions of the *csa* operon and homologous genes and operons can be obtained using similar methods.

csa operon homologous polynucleotide molecules can be isolated using standard hybridization techniques with probes of at least about 7 nucleotides in length and up to and including the full coding sequence. Homologous *csa* operon sequences can be identified using degenerate oligonucleotides capable of hybridization based on the sequences disclosed herein for use PCR amplification or by hybridization at moderate or

greater stringency. The term, "capable of hybridization" as used herein means that the subject nucleic acid molecules (whether DNA or RNA) anneal to an oligonucleotide of 15 or more contiguous nucleotides of SEQ. ID. NOs: 1, 3, 5, 7, and 9.

The choice of hybridization conditions will be evident to one skilled in the art and will generally be guided by the purpose of the hybridization, the type of hybridization (DNA-DNA or DNA-RNA), and the level of desired relatedness between the sequences. Methods for hybridization are well established in the literature. One of ordinary skill in the art realizes that the stability of nucleic acid duplexes will decrease with an increased number and location of mismatched bases; thus, the stringency of hybridization can be used to maximize or minimize the stability of such duplexes. Hybridization stringency can be altered by: adjusting the temperature of hybridization; adjusting the percentage of helix-destabilizing agents, such as formamide, in the hybridization mix; and adjusting the temperature and salt concentration of the wash solutions. In general, the stringency of hybridization is adjusted during the post-hybridization washes by varying the salt concentration and/or the temperature, resulting in progressively higher stringency conditions.

An example of progressively higher stringency conditions is as follows: 2 x SSC/0.1% SDS at about room temperature (hybridization conditions); 0.2 x SSC/0.1% SDS at about room temperature (low stringency conditions); 0.2 x SSC/0.1% SDS at about 42°C (moderate stringency conditions); and 0.1 x SSC at about 68°C (high stringency conditions). Washing can be carried out using only one of these conditions, e.g., high stringency conditions, or each of the conditions can be used, e.g., for 10-15 minutes each, in the order listed above, repeating any or all of the steps listed. As mentioned above, however, optimal conditions will vary, depending on the particular hybridization reaction involved, and can be determined empirically. In general, conditions of high stringency are used for the hybridization of the probe of interest.

Alternatively, polynucleotides having substantially the same nucleotide sequence set forth in SEQ. ID. NOs: 1, 3, 5, 7, and 9 or functional fragments thereof, or nucleotide sequences that are substantially identical to SEQ. ID. NOs: 1, 3, 5, 7, and 9, can represent members of a *csa*-like operon. By "substantially the same" or "substantially identical" is meant a nucleic acid or polypeptide exhibiting at least 80%,

85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% homology to a reference nucleic acid. For nucleotide sequences, the length of comparison sequences will generally be at least 10 to 500 nucleotides in length. More specifically, the length of comparison will be at least 50 nucleotides, at least 60 nucleotides, at least 75 nucleotides, and at least 110 nucleotides in length.

One embodiment of the invention provides isolated and purified polynucleotide molecules encoding one or more *csa* operon proteins, wherein the polynucleotide molecules that are capable of hybridizing under moderate to stringent conditions to an oligonucleotide of 15 or more contiguous nucleotides of SEQ. ID. NOs: 1, 3, 5, 7, and 9, including complementary strands thereto.

DNA sequences of the invention can be obtained by several methods. For example, the DNA can be isolated using hybridization or computer-based techniques which are well known in the art. Such techniques include, but are not limited to: 1) hybridization of genomic or cDNA libraries with probes to detect homologous nucleotide sequences; 2) antibody screening of expression libraries to detect cloned DNA fragments with shared structural features; 3) polymerase chain reaction (PCR) on genomic DNA or cDNA using primers capable of annealing to the DNA sequence of interest; 4) computer searches of sequence databases for similar sequences; and 5) differential screening of a subtracted DNA library.

Screening procedures which rely on nucleic acid hybridization make it possible to isolate any gene sequence from any organism, provided the appropriate probe is available. Oligonucleotide probes, which correspond to a part of *csa* operon sequences provided herein and encoding a Csa protein, can be synthesized chemically. This synthesis requires that short, oligo-peptide stretches of the amino acid sequence be known. The DNA sequence encoding the protein can be deduced from the genetic code, however, the degeneracy of the code must be taken into account. It is possible to perform a mixed addition reaction when the sequence is degenerate. This includes a heterogeneous mixture of denatured double-stranded DNA. For such screening, hybridization is preferably performed on either single-stranded DNA or denatured double-stranded DNA. Hybridization is particularly useful in the detection of cDNA clones derived from sources where an extremely low amount of mRNA sequences

relating to the polypeptide of interest are present. In other words, by using stringent hybridization conditions directed to avoid non-specific binding, it is possible, for example, to allow the autoradiographic visualization of a specific cDNA or DNA clone by the hybridization of the target DNA to that single probe in the mixture that is its complete complement. (Wallace, *et al., Nucl. Acid Res.*, 9:879, 1981). Alternatively, a subtractive library is useful for elimination of non-specific cDNA clones.

Among the standard procedures for isolating DNA sequences of interest is the formation of plasmid- or phage-carrying genomic libraries which include total DNA from the organism of interest. When used in combination with polymerase chain reaction technology, even rare expression products can be cloned. In those cases where significant portions of the amino acid sequence of the polypeptide are known, the production of labeled single or double-stranded DNA or RNA probe sequences duplicating a sequence putatively present in the target DNA can be employed in DNA/DNA hybridization procedures which are carried out on cloned copies of the DNA which have been denatured into a single-stranded form (Jay, *et al., Nucl. Acid Res.*, 11:2325, 1983).

The nucleotide sequences of the disclosed herein have a myriad of applications. Representative uses of the nucleotide sequences of the invention include the construction of DNA and oligonucleotide probes useful in Northern, Southern, and dot-blot assays for identifying and quantifying the level of expression of *csa* operon encoded proteins in a cell. *csa* operon encoded proteins have a variety of uses, for example, as antigens with which to elicit an immune response.

In addition, considering the important role the CS4 pili plays in ETEC attachment and colonization, it is thought highly likely that compositions containing the CS4 pili or compositions with activity against the same can result from expression of the *csa* operon. In this case, the proteins of the *csa* operon can prove highly useful in the generation of immunogenic compositions that can be used to generate an immune response in a subject.

Similarly, *csa* operon nucleotide sequences can be employed for the construction of recombinant cell lines, recombinant organisms, expression vectors, and the like. Such recombinant constructs can be used to express recombinant *csa* operon proteins.

In one embodiment, the recombinant constructs can be used to screen for candidate therapeutic agents capable of altering the pathology of a CS4 antigen-expressing organism. In another embodiment, the proteins of the *csa* operon present an attractive set of proteins with which to create immunogenic compositions. For example, the CS4 pili can be expressed, purified, and used to prepare an immunogenic subunit composition. In another embodiment, recombinant CS4 pili can be expressed in an organism, and the whole organism can be formulated into an immunogenic composition.

When the coding regions of the *csa* operon are used in the construction of various types of vectors, the *csa* sequences are often inserted into the coding region of the vector under the control of a promoter. Additionally, other elements, including regulatory elements, which are commonly found in vectors suitable for use in various molecular biology techniques, can also be included.

In one embodiment, a vector comprising a DNA molecule encoding a Csa protein is provided. Preferably, a DNA molecule including a *csaA*, *csaB*, *csaC*, *csaD*, or *csaE* gene, or a combination of these genes is inserted into a suitable expression vector, which is in turn used to transfect or transform a suitable host cell. Exemplary expression vectors include a promoter capable of directing the transcription of a polynucleotide molecule of interest in a host cell. Representative expression vectors include both plasmid and/or viral vector sequences. Suitable vectors include retroviral vectors, vaccinia viral vectors, CMV viral vectors, BLUESCRIPT (Stratagene, San Diego, CA) vectors, bacculovirus vectors, and the like. In another embodiment, promoters capable of directing the transcription of a cloned gene or cDNA can be inducible or constitutive promoters and include viral and cellular promoters.

In some embodiments, it can be preferable to use a selectable marker to identify cells that contain the cloned DNA. Selectable markers are generally introduced into the cells along with the cloned DNA molecules and include genes that confer resistance to drugs, such as ampicillin, neomycin, hygromycin, and methotrexate. Selectable markers can also complement auxotrophies in the host cell. Other selectable markers provide detectable signals, such as beta-galactosidase to identify cells containing the cloned DNA molecules.

Antisense

Antisense *csa* nucleotide sequences can be used to block *csa* expression. Suitable antisense oligonucleotides are at least 11 nucleotides in length and can include untranslated (upstream) and associated coding sequences. As will be evident to one skilled in the art, the optimal length of an antisense oligonucleotide depends on the strength of the interaction between the antisense oligonucleotide and the complementary mRNA, the temperature and ionic environment in which translation takes place, the base sequence of the antisense oligonucleotide, and the presence of secondary and tertiary structure in the mRNA and/or in the antisense oligonucleotide. Suitable target sequences for antisense oligonucleotides include promoter regions, ribosome binding sites, and sites that interfere with ribosome progression.

Antisense oligonucleotides can be prepared, for example, by the insertion of a DNA molecule containing the target DNA sequence into a suitable expression vector such that the DNA molecule is inserted downstream of a promoter in a reverse orientation as compared to the particular *csa* gene itself. The expression vector can then be transduced, transformed or transfected into a suitable cell resulting in the expression of antisense oligonucleotides. Alternatively, antisense oligonucleotides can be synthesized using standard manual or automated synthesis techniques. Synthesized oligonucleotides are introduced into suitable cells by a variety of means including electroporation, calcium phosphate precipitation, or microinjection. The selection of a suitable antisense oligonucleotide administration method will be evident to one skilled in the art.

With respect to synthesized oligonucleotides, the stability of antisense oligonucleotide-mRNA hybrids is advantageously increased by the addition of stabilizing agents to the oligonucleotide. Stabilizing agents include intercalating agents that are covalently attached to either or both ends of the oligonucleotide. In preferred embodiments, the oligonucleotides are made resistant to nucleases by, for example, modifications to the phosphodiester backbone by the introduction of phosphotriesters, phosphonates, phosphorothioates, phosphoroselenoates, phosphoramidates, phosphorodithioates, or morpholino rings.

Amino Acids

The identification of the csa operon and the proteins encoded thereby is described. Representative polypeptides produced from the genes of the *csa* operon include sequences comprising CsaA (SEQ. ID. NO.: 2), CsaB (SEQ. ID. NO.: 4), CsaC (SEQ. ID. NO.: 6), CsaD (SEQ. ID. NO.: 8), and CsaE (SEQ. ID. NO.: 10). Variants of the *csa* operon encoded proteins include those amino acid sequences resulting from in minor genetic polymorphisms, differences between strains, and those that contain amino acid substitutions, additions, and/or deletions.

The *csa* operon encoded proteins, as described herein, encompass the whole proteins encoded by the operon, as well as fragments of *csa* proteins that are functionally active. *csa* operon encoded proteins purified from naturally occurring materials and closely related, functionally similar proteins retrieved by antisera specific to the *csa* proteins, and recombinantly expressed proteins encoded by genetic materials (DNA, RNA, cDNA) retrieved on the basis of their similarity to regions in the *csa* operon sequences are also encompassed by the present description.

According to the present description, polynucleotide molecules encoding *csa* operon encoded proteins encompass those molecules that encode *csa* proteins or peptides that share identity with the sequences shown in SEQ. ID. NOs.: 2, 4, 6, 8, and 10. Such molecules preferably share greater than 30% identity at the amino acid level with the disclosed sequences in *csa*. In preferred embodiments, the polynucleotide molecules can share greater identity at the amino acid level across highly conserved regions.

It is contemplated that amino acid sequences substantially the same as the sequences set forth in SEQ. ID. NOs.: 2, 4, 6, 8, and 10, are encompassed by the present description. A preferred embodiment includes polypeptides having substantially the same sequence of amino acids as the amino acid sequence set forth in SEQ ID NOs.: 2, 4, 6, 8, and 10, or functional fragments thereof, or amino acid sequences that are substantially identical to SEQ ID NOs.: 2, 4, 6, 8, and 10. By "substantially the same" or "substantially identical" is meant a polypeptide exhibiting at least 80%, 85%, 90%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% homology to a

reference amino acid sequence. For polypeptides, the length of comparison sequences will generally be at least 16 amino acids, preferably at least 20 amino acids, more preferably at least 25 amino acids, and most preferably 35 amino acids.

Homology is often measured using sequence analysis software (*e.g.*, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705 or the NCBI BLAST program). Such software matches similar sequences by assigning degrees of homology to various substitutions, deletions, substitutions, and other modifications.

The term "functional fragments" include those fragments of SEQ ID NOs.: 2, 4, 6, 8, and 10, or other proteins that have a similar amino acid sequence as that of the *csa* operon encoded proteins, that retain the function or activity of the various *csa* proteins. One of skill in the art can screen for the functionality of a fragment by using the examples provided herein, where full-length *csa* operon encoded proteins are described. It is also envisioned that fragments of various *csa* operon encoded proteins can be identified in a similar manner.

By "substantially identical" is also meant an amino acid sequence which differs only by conservative amino acid substitutions, for example, substitution of one amino acid for another of the same class (*e.g.*, valine for glycine, arginine for lysine, *etc.*) or by one or more non-conservative substitutions, deletions, or insertions located at positions of the amino acid sequence which do not destroy the function of the protein assayed, (*e.g.*, as described herein). Preferably, such a sequence is at least 85%, and more preferably from 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, to 100% homologous at the amino acid level to SEQ ID NOs:2, 4, 6, 8, or 10.

By a "substantially pure polypeptide" is meant a *csa* operon encoded protein that has been separated from components that naturally accompany it. Typically, the polypeptide is substantially pure when it is at least 60%, by weight, free from the proteins and other naturally occurring molecules with which it is typically associated. Preferably, the preparation is at least 75%, 80%, 90%, 95%, and most preferably at least 99%, by weight, *csa* operon encoded protein. A substantially pure *csa* operon encoded polypeptide can be obtained, for example, by extraction from a natural source; by expression of a recombinant nucleic acid encoding a *csa* operon encoded polypeptide;

or by chemically synthesizing the protein. Purity can be measured by any appropriate method, *e.g.*, column chromatography, polyacrylamide gel electrophoresis, or by HPLC analysis.

A protein is substantially free of naturally associated components when it is separated from those contaminants that accompany it in its natural state. Thus, a protein that is chemically synthesized or produced in a cellular system different from the cell from which it naturally originates will be substantially free from its naturally associated components. Accordingly, substantially pure polypeptides include those derived from eukaryotic organisms but synthesized in *E. coli* or other prokaryotes.

As would be evident to one skilled in the art, the polynucleotide molecules of the present disclosure can be expressed in a variety of prokaryotic and eucaryotic cells using regulatory sequences, vectors, and methods well established in the literature.

csa operon encoded proteins produced according to the present description can be purified using a number of established methods such as affinity chromatography using an anti-Csa protein antibodies coupled to a solid support. Fusion proteins of an antigenic tag and a *csa* operon encoded protein can be purified using antibodies to the tag. Optionally, additional purification is achieved using conventional purification means such as liquid chromatography, gradient centrifugation, and gel electrophoresis, among others. Methods of protein purification are known in the art and can be applied to the purification of recombinant *csa* operon encoded proteins described herein. Purification of *csa* operon products is discussed more completely below.

Construction of *csa* operon encoded fusion proteins is also contemplated. Fusion proteins will typically contain additions, substitutions, or replacements of one or more contiguous amino acids of the native *csa* operon encoded protein with amino acid(s) from a suitable fusion protein partner. Such fusion proteins are obtained using recombinant DNA techniques well known by one of skill in the art. Briefly, DNA molecules encoding the hybrid *csa* operon encoded proteins of interest are prepared using generally available methods such as PCR mutagenesis, site-directed mutagenesis, and/or restriction digestion and ligation. The hybrid DNA is then inserted into expression vectors and introduced into suitable host cells.

One embodiment of the present invention involves the isolation of proteins that interact with *csa* operon encoded proteins or are receptors for CS4 pili. *csa* operon encoded proteins can be used in immunoprecipitation to isolate interacting factors or used for the screening of interactors using different methods of two hybrid screening. Isolated interactors of *csa* operon encoded proteins can be used to modify or block CS4 mediated binding to a host cell.

Synthetic peptides, recombinantly derived peptides, fusion proteins, chiral proteins (stereochemical isomers, racemates, enantiomers, and D-isomers) and the like are provided which include a portion of a *csa* operon encoded protein or the entire protein. The subject peptides have an amino acid sequence encoded by a nucleic acid which hybridizes under stringent conditions with an oligonucleotide of 15 or more contiguous nucleotides of SEQ. ID. NOs: 1, 3, 5, 7, and 9. Representative amino acid sequences of the subject peptides are disclosed in SEQ. ID. NOs: 2, 4, 6, 8, and 10. The subject peptides find a variety of uses, including preparation of specific antibodies and preparation of antagonists of CS4 binding.

Antibodies

As noted above, the described teachings provide antibodies that bind to *csa* operon encoded proteins. The production of non-human antisera or monoclonal antibodies (e.g., murine, lagomorph, porcine, equine) is well known and can be accomplished by, for example, immunizing an animal with a *csa* operon encoded protein or peptides. For the production of monoclonal antibodies, antibody producing cells are obtained from immunized animals, immortalized and screened, or screened first for the production of the antibody that binds to the particular *csa* operon encoded protein or peptides and then immortalized. It can be desirable to transfer the antigen binding regions (e.g., F(ab')2 or hypervariable regions) of non-human antibodies into the framework of a human antibody by recombinant DNA techniques to produce a substantially human molecule.

Following synthesis or expression and isolation or purification of a *csa* operon encoded protein or a portion thereof, the isolated or purified protein can be used to generate antibodies and tools for identifying agents that interact with the *csa* operon

encoded protein and fragments of interest. Depending on the context, the term "antibodies" can encompass polyclonal, monoclonal, chimeric, single chain, Fab fragments and fragments produced by a Fab expression library. Antibodies that recognize *csa* operon encoded proteins and fragments thereof have many uses including, but not limited to, biotechnological applications, therapeutic/prophylactic applications, and diagnostic applications.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, etc., can be immunized by injection with *csa* operon encoded proteins or any portion, fragment or oligopeptide that retains immunogenic properties. Depending on the host species, various adjuvants can be used to increase immunological response. Such adjuvants include, but are not limited to, detoxified heat labile toxin from *E. coli*, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. BCG (*Bacillus Calmette-Guerin*) and *Corynebacterium parvum* are also potentially useful adjuvants.

Peptides used to induce specific antibodies can have an amino acid sequence consisting of at least three amino acids, and preferably at least 10 to 15 amino acids. Preferably, short stretches of amino acids encoding fragments of *csa* operon encoded proteins are fused with those of another protein such as keyhole limpet hemocyanin such that an antibody is produced against the chimeric molecule. While antibodies capable of specifically recognizing *csa* operon encoded proteins can be generated by injecting synthetic 3-mer, 10-mer, and 15-mer peptides that correspond to a protein sequence of the *csa* operon encoded protein or proteins of interest into mice, a more diverse set of antibodies can be generated by using recombinant *csa* operon encoded proteins, purified *csa* operon encoded proteins, or fragments of *csa* operon encoded proteins.

To generate antibodies to *csa* operon encoded proteins and fragments thereof, a substantially pure *csa* operon encoded protein or a fragment thereof is isolated from a transfected or transformed cell or the wildtype ETEC organism. The concentration of the polypeptide in the final preparation is adjusted, for example, by concentration on an

Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the polypeptide of interest can then be prepared as follows:

Monoclonal antibodies to *csa* operon encoded proteins or a fragment thereof can be prepared using any technique that provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique originally described by Koehler and Milstein (*Nature* 256:495-497 (1975), the human B-cell hybridoma technique (Kosbor et al. *Immunol Today* 4:72 (1983); Cote et al *Proc Natl. Acad. Sci* 80:2026-2030 (1983), and the EBV-hybridoma technique Cole et al. Monoclonal Antibodies and Cancer Therapy, Alan R. Liss Inc, New York N.Y., pp 77-96 (1985). In addition, techniques developed for the production of "chimeric antibodies", the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can be used. (Morrison et al. *Proc Natl. Acad. Sci* 81:6851-6855 (1984); Neuberger et al. Nature 312:604-608(1984); Takeda et al. *Nature* 314:452-454(1985). Alternatively, techniques described for the production of single chain antibodies (U.S. Pat. No. 4,946,778) can be adapted to produce *csa* operon encoded protein-specific single chain antibodies. Antibodies can also be produced by inducing *in vivo* production in the lymphocyte population or by screening recombinant immunoglobulin libraries or panels of highly specific binding reagents as disclosed in Orlandi et al., *Proc Natl. Acad. Sci* 86: 3833-3837 (1989), and Winter G. and Milstein C; *Nature* 349:293-299 (1991).

Antibody fragments that contain specific binding sites for *csa* operon encoded proteins can also be generated. For example, such fragments include, but are not limited to, the $F(ab')_2$ fragments that can be produced by pepsin digestion of the antibody molecule and the Fab fragments that can be generated by reducing the disulfide bridges of the $F(ab')_2$ fragments. Alternatively, Fab expression libraries can be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (Huse W. D. et al. *Science* 256:1275-1281 (1989)).

By one approach, monoclonal antibodies to *csa* operon encoded proteins or fragments thereof are made as follows. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein or peptides derived therefrom, over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen

isolated. The spleen cells are fused in the presence of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, E., *Meth. Enzymol.* 70:419 (1980), and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. et al. Basic Methods in Molecular Biology Elsevier, New York. Section 21-2.

Polyclonal antiserum containing antibodies to heterogeneous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein or peptides derived therefrom described above, which can be unmodified or modified to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than others and can require the use of carriers and adjuvant. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. et al. *J. Clin. Endocrinol. Metab.* 33:988-991 (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. et al., Chap. 19 in: Handbook of Experimental Immunology D. Wier (ed) Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12μM). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: Manual of Clinical Immunology, 2d Ed. (Rose and Friedman, Eds.) Amer. Soc. For Microbiol., Washington, D.C. (1980). Antibody preparations prepared

according to either protocol are useful in quantitative immunoassays that determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively (e.g., in diagnostic embodiments that identify the presence of a *csa* operon encoded protein in biological samples). It is also contemplated that various methods of molecular modeling and rational drug design can be applied to identify compounds that resemble a *csa* operon encoded protein, fragment, or derivative thereof, and molecules that interact with *csa* operon encoded proteins and, thereby modulate their function.

ExpressionVectors

Recombinant gene expression vectors comprising the *csa* operon, or portions thereof, can be constructed in a variety of forms well-known in the art. Preferred expression vectors include plasmids and cosmids. Expression vectors include one or more fragments of the *csa* operon. Typically, an expression vector will comprise one or more genes of *csa* operon. In one embodiment, an expression vector will comprise and operatively encode the *csaA*, *csaB*, *csaC*, *csaE*, and *csaD* coding regions. Alternative embodiments of the described expression vectors can have various combinations of the coding regions *csa* operon. For example, an expression can comprise the *csaB*, the *csaE* coding regions, or a combination of both.

As used herein, the phrase "operatively encode" refers to one or more protein coding regions associated with those regulatory sequences required for expression of the polypeptide encoded by the coding region. Examples of such regulatory regions including promoter binding sites, enhancer elements, ribosome binding sites, and the like. Those of ordinary skill in the art will be able to select regulatory sequences and incorporate them into the recombinant expression vectors described herein without undue experimentation. For example, suitable regulatory sequences for use in various eukaryotic and prokaryotic systems are described in Ausubel, et al., Short Protocols in Molecular Biology, 3rd ed., John Wiley & Sons, Inc, New York, 1997, which is hereby incorporated by reference in its entirety.

Expression vectors for use with the *csa* operon will typically contain regulatory sequences derived from a compatible species for expression in the desired host cell. For

example, when *E. coli* is the host cell, the host cell population can be typically transformed using pBR322, a plasmid derived from an *E. coli* species. (Bolivar, et al., Gene, 2:95, 1977). pBR322 contains genes for ampicillin (AMPR) and tetracycline resistance and thus provides easy means for identifying transformed cells.

Promoters suitable for use with prokaryotic hosts illustratively include the beta-lactamase and lactose promoter systems (Chang, et al., Nature, 275:615, 1978; and Goeddel, et al., Nature, 281:544, 1979), alkaline phosphatase, the tryptophan (trp) promoter system (Goeddel, Nucleic Acids Res., 8:4057, 1980) and hybrid promoters such as the *taq* promoter (de Boer, et al., Proc. Natl. Acad. Sci. USA, 80:21-25, 1983). Other functional bacterial promoters are also suitable. Their nucleotide sequences are generally known in the art, thereby enabling a skilled worker to ligate them to a polynucleotide which encodes the peptide of interest (Siebenlist, et al., Cell, 20:269, 1980) using linkers or adapters to supply any required restriction sites.

In addition to prokaryotes, eukaryotic microbes such as yeast cultures can also be used as source for the regulatory sequences. *Saccharomyces cerevisiae* is a commonly used eukaryotic host microorganism. Suitable promoting sequences for use with yeast hosts include the promoters for 3-phosphoglycerate kinase (Hitzeman, et al., J. Biol. Chem., 255:2073, 1980) or other glycolytic enzymes (Hess, et al. J. Adv. Enzyme Reg. 7:149, 1968; and Holland, Biochemistry, 17:4900, 1978) such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase.

Other yeast promoters, which are inducible promoters having the additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degraded enzymes associated with nitrogen metabolism, metallothionine, glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Yeast enhancers also are advantageously used with yeast promoters.

In another embodiment, a recombinant virus is used as the expression vector. Exemplary viruses include the adenoviruses, adeno-associated viruses, herpes viruses,

vaccinia, or an RNA virus such as a retrovirus or an alphavirus. Preferably, the retroviral vector is a derivative of a murine or avian retrovirus. Preferably the alphavirus vector is derived from Sindbis or Semliki Forest Virus. All of these vectors can transfer or incorporate a gene for a selectable marker so that transduced cells can be identified and generated.

By inserting one or more sequences of interest into the viral vector, along with another gene which encodes the ligand for a receptor on a specific target cell, for example, the vector is now target specific. Retroviral vectors can be made target specific by inserting, for example, a polynucleotide encoding a sugar, a glycolipid, or a protein. Preferred targeting is accomplished by using an antibody to target the retroviral vector, such as to the vicinity of a mucosal inductor site, using a MALT-specific antibody. Those of skill in the art will know of, or can readily ascertain without undue experimentation, specific polynucleotide sequences which can be inserted into the retroviral genome to allow target specific delivery of the retroviral vector containing the polynucleotides of interest.

It will be appreciated that the same techniques that are utilized to incorporate the nucleotide sequences of the *csa* operon and optionally other immunostimulatory polynucleotides into viral gene expression vectors can be used to incorporate the sequences into live and attenuated live viruses for use as immunogenic compositions.

Targeting of mucosal tissues can be performed by exploiting inherent biological properties of the lymphoid bed which is to be targeted. These include the crypt architecture of the tonsillar pillars which can be used to entrap particles, and also include the M cells of Peyer's patches in the gut, which M cells specifically endocytose a wide variety of particles including lipid particles and other small particulates. Therefore, those skilled in the art can prepare a wide variety of molecular particulate preparations which, if provided to intestine, will lodge within the crypt portions of intestinal Peyer's patches and be endocytosed by M cells. If such particles provide for delivery of a biologically active polynucleotide to M cells, then such particles will enable the stimulation or modulation of mucosal immune response induction by the Peyer's patch lymphoid tissue to which the M cell traffics.

Construction of suitable vectors containing desired coding, non-coding and control sequences employ standard ligation techniques. Isolated plasmids or DNA fragments are cleaved, tailored, and re-ligated in the form desired to construct the plasmids required.

For example, for analysis to confirm correct sequences in plasmids constructed, the ligation mixtures can be used to transform a host cell and successful transformants selected by antibiotic resistance where appropriate. Plasmids from the transformants are prepared, analyzed by restriction and/or sequenced by, for example, the method of Messing, et al., (Nucleic Acids Res., 9:309, 1981), the method of Maxam, et al., (Methods in Enzymology, 65:499, 1980), or other suitable methods which will be known to those skilled in the art. Size separation of cleaved fragments is performed using conventional gel electrophoresis as described, for example, by Maniatis, et al., (Molecular Cloning, pp. 133-134, 1982).

Host cells can be transformed with the expression vectors described herein and cultured in conventional nutrient media modified as is appropriate for inducing promoters, selecting transformants or amplifying genes. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

Purification

The purification of one or more of the products of the *csa* operon is disclosed below. Steps involved in the purification of *csa* operon products include (1) solubilization of the desired protein, (2) the development of one or more isolation and concentration procedures, (3) stabilization of the protein following purification, and (4) development of a suitable assay to determine the presence of the desired protein. Various aspects of protein isolation and purification are discussed in detail in Cooper, T. G., "The Tools of Biochemistry," John Wiley & Sons, New York, 1977, which is hereby incorporated by reference in its entirety. As the techniques of protein isolation and purification are notoriously well known in the art, this disclosure will refrain from discussing them in detail. Nevertheless, elements of the cited reference are summarized and discussed below.

Solubilization is required of most proteins to be purified, as most isolation procedures commonly used operate in aqueous solutions. In some cases solubilization can be achieved by merely lysing a host cell within which a desired protein has been expressed. In other situations, additional steps, such as extracting the desired protein from a subcellular organelle may be required. Osmotic lysis, grinding, the use of blenders, ultrasonic waves, presses, and other well known techniques of protein solubilization are contemplated for use with the methods disclosed herein.

Regarding the isolation and concentration of *csa* operon products, there are variety of techniques available that are well known in the art. These techniques include (1) differential solubility, (2) ion exchange chromatography, (3) absorption chromatography, (4) molecular sieve techniques, (5) affinity chromatography, (6) electrophoresis, and (7) electrofocusing. Each of these techniques can be useful in the purification of one or more *csa* operon products. An immunogenic purification methodology for the purification of the CS4 pili is described in Wolf, et al., "Characterization of CS4 and CS6 antigenic components of PCF8775, a putative colonization factor complex from enterotoxigenic *Escherichia coli* E8775," *Infect Immun.* 57(1):164-73 (1989), which is hereby incorporated by reference in its entirety. For purifying intact CS4 fimbriae, use the methods described in Hall, et al., *J. Bacteriol.* 171:6372 (1989).

Stabilizing and maintaining a purified product of the *csa* operon in a functional state warrants attention to a number of different conditions. These conditions include (1) pH, (2) degree of oxidation, (3) heavy metal concentration, (4) medium polarity, (5) protease concentration, and (6) temperature. One of ordinary skill in the art would readily know which of the available techniques to use to maintain purified *csa* operon products in an active form without undue experimentation.

Developing one or more assays with which to determine the presence and functionality of the purified product of the *csa* operon will hinge on the individual proteins themselves. Perhaps the most useful assay to develop involves the generation of one or more antibodies with which to identify the various products of the *csa* operon. The generation of antibodies is discussed herein. Because the nucleotide and amino acid sequences of each of the csa operon encoded gene sequences is known, it would be

trivial for one of ordinary skill in the art to generate suitable antibodies for the detect of Csa proteins, using techniques that are notoriously well known.

Compositions

The proteins encoded by the *csa* operon can be used to formulate immunogenic compositions that facilitate an immune response. Examples of a typical immune response include a mucosal immune response and a systemic immune response. A variety of embodiments utilizing the proteins of the *csa* operon are envisioned.

Recombinant organisms

In one embodiment, a nucleotide sequence comprising the *csa* operon or a functional fragment thereof is introduced into an exogenous organism using standard molecular biology techniques well known to those of ordinary skill in the art. Exemplary techniques are discussed in Ausubel, et al., "Short Protocols in Molecular Biology." The resulting recombinant organism can then be used as an immunogen against which an immune response may be engendered. In a preferred embodiment, an attenuated pathogenic organism serves as the exogenous organism. It is contemplated that an entire recombinant organism or a functional fragment thereof, such as an isolated membrane fraction, liposome, or the like, can be used to generate an immunogenic composition.

For example, one application of the discoveries described herein is directed to the development of a multivalent hybrid vaccine to prevent both *Shigella* dysentery and ETEC diarrhea, (Altboum, Z., et al., Attenuated *Shigella flexneri* 2a Δ*guaBA* strain CVD 1204 expressing ETEC CS2 and CS3 fimbriae as a live mucosal vaccine against *Shigella* and enterotoxigenic *Escherichia coli* infection, In press; Koprowski, et al., *Infect. Immun.*, 68:4884-92 (2000); Kotloff, et al., *Vaccine*, 13:495-502 (1995); Levine, et al., "Fimbrial vaccines," *In* P. Klemm (ed.), Fimbriae: adhesion, biogenics, genetics and vaccines, Boca Raton: CRC Press, 1994; Noriega, et al., *Infect. Immun.*, 64:23-27 (1996)). United States Patent No. 6,190,669, to Noriega, et al., entitled "Attenuated mutants of salmonella which constitutively express the Vi antigen," which is hereby incorporated by reference in its entirety, contains additional teaching relating to the

generation of such chimeric organisms suitable for use in the preparation of an immunogenic composition containing the CS4 pili.

In a preferred embodiment, the ETEC-CS4 fimbrial encoding genes of the *csa* operon are isolated and used to construct an efficient multivalent *Shigella*- ETEC immunogenic composition that will protect from diarrhea caused by either *Shigella* and CS4 expressing ETEC strains. One aspect of this embodiment is directed to creating an immunogenic composition capable of generating an immune response against LTh, the LT variant found in human ETEC strains. The immunogenic properties of cloned CFA/I, CS2, CS3 and LThK63 encoding genes in *S. flexneri* 2a strain CVD 1204 have been reported. This work has been expanded by including the expression of the CS4 pili as an intact fimbriae in both *E.coli* and *Shigella* strains.

Other multivalent immunogenic compositions effective against enteric bacteria are contemplated. For example, a multivalent immunogenic composition against *Salmonella* spp.; *Clostridium* spp., such as *Clostridium botulinum*; *Staphylococcus* spp , such as *S. aureus*; *Campylobacter* spp., such as *C. jejuni*; *Yersinia* spp., such as *Y. enterocolitica* and *Y. pseudotuberculosis*; *Listeria* spp., such as *L. monocytogenes*; various *Vibrio* spp., including *V. cholerae* O1, *V. cholerae* non-O1, *V. parahaemolyticus*, *V. vulnificus*; *Clostridium* spp., such as *C. perfringens*; *Bacillus* spp., such as *B. cereus*; *Aeromonas* spp., such as *A. hydrophila*; *Plesiomonas* spp., such as *P. shigelloides*; *Shigella* spp.; *Streptococcus* spp.; various miscellaneous enterics such as *Klebsiella* spp.; *Enterobacter* spp.; *Proteus* spp.; *Citrobacter* spp.; *Aerobacter* spp.; *Providencia* spp.; *Serratia* spp.; and members of the enterovirulent *Escherichia coli* Group (EEC Group) which comprises, enterotoxigenic *Escherichia coli* (ETEC), enteropathogenic *Escherichia coli* (EPEC), enterohemorrhagic *Escherichia coli* (EHEC) such as *Escherichia coli* O157:H7, and enteroinvasive *Escherichia coli* (EIEC).

Subunit vaccines

Another embodiment described herein relates to the generation of immunogenic compositions comprising distinct immunogenic proteins or fragments thereof, or functional fragments of organisms of interest. Such immunogenic compositions are referred to here as subunit immunogenic composition because at least one of the

components of the composition is a subunit of an organism, rather than an entire organism. Typically, a subunit immunogenic composition as described herein comprises one or more immunogenic components.

In a preferred embodiment, the subunit immunogenic composition described herein comprise a carrier component and an immunogenic component. Typically, the carrier component will function as a binding moiety with which the originating organism uses to bind to and gain entrance into the host organism. One example of such a carrier component is the CS4 antigen, which is encoded by the *csa* operon. The CS4 antigen allows ETEC strains expressing this protein complex to bind to the mucosa of a human host. In a preferred embodiment the CsaB and/or the CsaE proteins can function as carrier components.

Although not wishing to be bound by theory, it is hypothesized that the subunit immunogenic composition described herein function by exposing the immunogenic component of the subunit immunogenic composition to the mucosa, and various immune system components present there. In one theory, the generation of a desired immune response by the subunit immunogenic compositions described herein occurs by increasing the exposure of the immunogenic compositions to the target tissue. The presence of both a carrier component and an immunogenic component are theorized to achieve this goal.

Any protein, peptide, or amino acid sequence that elicits an immune response can be used as the immunogenic component in the subunit immunogenic compositions described herein.

The carrier components of the described subunit immunogenic compositions can also possess immunogenic characteristics themselves. Typically, adjuvants are used in immunogenic compositions to enhance the immune response directed against the immunogenic component of the immunogenic compositions disclosed. Carrier components that possess both mucosa binding characteristics and immunogenic characteristics can be used. For example, in one embodiment, the CsaB and/or the CsaE proteins can function both as the carrier component and the immunogenic component.

For the carrier components described above, the entire molecule can be used as the carrier component, or a functionally active fragment of the molecule can be used. Mutagenized forms of these molecules can also be used as carrier components.

In one embodiment, one or more Csa proteins are isolated and purified. The one or more Csa proteins can be mixed or coupled with one or more immunogenic compounds. For example, in one embodiment, CsaB can be expressed, purified, and cross-linked to a toxin, toxoid (an attenuated toxin), or some other immunogenic compound for use in a subunit immunogenic compositions.

In another embodiment, expressed Csa proteins or fragments thereof, or whole CS4 antigen complexes can be cross-linked to an immunogenic component. The immunogenic component can also be an isolated protein, functional fragment thereof, whole organism (such as a bacterium or a virus), or functional fragment thereof, that is isolated either in part or as is used as a whole pathogenic organism. The Csa proteins can be isolated from the bacterium itself or it can be produced using recombinant DNA techniques well known in the art.

In accordance with one aspect of the present invention, the smaller fragments of expression product of the *csa* operon are used to provide an immunogenic composition. Specifically, these fragments will comprise an immunogenic region of such expression product, typically from about 5, 6, 8, 10 or 12 amino acids to about 20, 22, 24, 30, or more amino acids. Suitable fragments or immunogenic regions can be readily ascertained using as screening procedures the techniques set forth in Examples 1-4. In one embodiment of a suitable screening procedure, a large number of candidate fragments are more or less randomly produced and used to immunize guinea pigs or other suitable models. Alternatively, full-length polypeptides shown to be active in the present invention can be truncated and screened in an iterative process to isolate the immunogenic and protective activity to a minimal fragment. Such screening can be readily carried out without undue experimentation and the active fragments are within the contemplation of the present invention.

The Csa proteins used to form the immunogenic compositions of this embodiment can be the whole protein, such as the CsaB or CsaE proteins, an immunogenic fragment of a Csa protein, a mutagenized form of a Csa protein, or a

fusion protein comprising a Csa protein or a fragment thereof and a suitable fusion partner. A suitable fusion partner for such a Csa fusion protein or a fragment thereof might be any protein, peptide or amino acid sequence that facilitates the expression and/or purification of the Csa fusion protein using recombinant DNA techniques known in the art. Alternatively, one or more additional immunogens can serve as the fusion partner in a Csa fusion protein.

Nucleotide Immunogenic Compositions

In another embodiment, an immune response can be elicited using nucleotide-containing compositions. In one aspect of this embodiment, a mucosal or systemic immune response is elicited in a host by administering an antigen-encoding polynucleotide preparation, comprising DNA or RNA, which encodes an antigenic epitope to the host. In a preferred embodiment, the nucleotide-containing composition is administered to a mucosal inductor site in the mucosal tissue of the host.

Naked DNA may be administered directly to mucosa, for instance in saline drops, or in a recombinant gene expression vector. Preferably, the recombinant gene expression vectors are not capable of replication or dissemination. Nucleotide-containing immunogenic compositions also comprise live viral immunogenic compositions wherein the viruses include immunostimulatory polynucleotides. According to a preferred method of the invention, a target protein antigen is administered through its expression by a recombinant gene expression vector.

U.S. Patent No. 6,110,898, to Malone, et al., entitled, "DNA vaccines for eliciting a mucosal immune response," which is hereby incorporated by reference in its entirety, provides detailed teaching for the generation of such immunogenic compositions.

Formulations and Administration

The immunogenic compositions described herein can be formulated in a variety of useful formats for administration by a variety of routes. Concentrations of the immunogenic components in the formulations described will be such that an effective

dose of the immunogenic components are included in the formulation. Determination of such a concentration would be readily apparent to those of ordinary skill in the art.

Administration of the immunogenic compositions can be by nasal application, by inhalation, ophthalmically, orally, rectally, vaginally, or by any other mode that results in the immunogenic contacting the mucosal tissues.

In one embodiment, the immunogenic composition exists as an atomized dispersion for use in delivery by inhalation. The atomized dispersion of the immunogenic components will typically contain carriers for atomized or aerosolized dispersions, such as buffered saline and other compounds well known to those of skill in the art. The delivery of the described immunogenic compositions via inhalation has the effect of rapid dispersion to a large area of mucosal tissues, as well as absorption by the blood for circulation of the immunogenic components. One example of a method of preparing an atomized dispersion is found in U.S. Patent 6,187,344, entitled, "Powdered Pharmaceutical Formulations Having Improved Dispersibility," which is hereby incorporated by reference in its entirety.

The immunogenic compositions described herein can also be formulated in the form of a suppository, whether rectal or vaginal. Typical carriers for formulation of the inactive portion of a suppository include polyethylene glycol, glycerine, cocoa butter and other compounds well known to those of skill in the art. Other suppository formulations suitable for delivery of the described immunogenic compositions are also contemplated. Delivery of the described immunogenic compositions via suppository it hypothesized to have the effect of contacting a mucosal surface with the immunogenic compositions for release to proximal mucosal tissues. Distal mucosal tissues also receive the immunogenic compositions by diffusion.

Additionally, immunogenic compositions are contemplated as existing in a liquid form. The liquid can be for oral dosage, for ophthalmic or nasal dosage as drops, or for use as an enema or douche. When the immunogenic compositions is formulated as a liquid, the liquid can be either a solution or a suspension of the immunogenic compositions.

A colloidal dispersion system may be used for targeted delivery of nucleic acid-containing immunogenic compositions. Colloidal dispersion systems include

macromolecule complexes, nanocapsules, microspheres, beads, and lipid-based systems including oil-in-water emulsions, micelles, mixed micelles, and liposomes. A preferred colloidal system is a lipid preparation including unilamaller and multilamellar liposomes.

Liposomes are artificial membrane vesicles that are useful as delivery vehicles in vitro and in vivo. It has been shown that large unilamellar vesicles (LUV), which range in size from 0.2-4.0 μm can encapsulate a substantial percentage of an aqueous buffer containing large macromolecules. RNA, DNA and intact virions can be encapsulated within the aqueous interior and be delivered to cells in a biologically active form (Fraley, et al., Trends Biochem. Sci., 6:77, 1981). In addition to mammalian cells, liposomes have been used for delivery of polynucleotides in plant, yeast and bacterial cells. In order for a liposome to be an efficient gene transfer vehicle, the following characteristics should be present: (1) encapsulation of the genes encoding the polynucleotides at high efficiency while not compromising their biological activity; (2) preferential and substantial binding to a target cell in comparison to non-target cells; (3) delivery of the aqueous contents of the vesicle to the target cell cytoplasm at high efficiency; and (4) accurate and effective expression of genetic information (Mannino, et al., Biotechniques, 6:682, 1988). In addition to such LUV structures, multilamellar and small unilamellar lipid preparations that incorporate various cationic lipid amphiphiles can also be mixed with anionic polynucleotides to form nucleolipidic particles which are often also referred to as liposomes (Felgner, et al, Proc Natl. Acad. Sci. U.S.A. 84 (21): 7413 1987) and used to deliver the nucleic acids into cells.

The composition of the liposome is usually a combination of phospholipids, particularly high-phase-transition-temperature phospholipids, usually in combination with steroids, especially cholesterol. Other phospholipids or other lipids may also be used. The physical characteristics of liposomes depend on pH, ionic strength, and the presence of divalent cations. The appropriate composition and preparation of cationic lipid amphiphile:polynucleotide formulations are known to those skilled in the art, and a number of references which provide this information are available (*e.g.*, Bennett, et al, J. Liposome Res. 6(3):545).

Examples of lipids useful in liposome production include phosphatidyl compounds, such as phosphatidylglycerol, phosphatidylcholine, phosphatidylserine, phosphatidylethanolamine, sphingolipids, cerebrosides, and gangliosides. Particularly useful are diacylphosphatidylglycerols, where the lipid moiety contains from 14-18 carbon atoms, particularly from 16-18 carbon atoms, and is saturated. Illustrative phospholipids include egg phosphatidylcholine, dipalmitoylphosphatidylcholine and distearoylphosphatidylcholine. Examples of cationic amphiphilic lipids useful in formulation of nucleolipid particles for polynucleotide delivery include the monovalent lipids DOTAP, DOTMA, and DC-Chol, the polyvalent lipids LipofectAMINE, DOGS, Transfectam and other amphiphilic polyamines. These agents may be prepared with helper lipids (such as Dioleoyl Phosphatidyl Ethanolamine) or with various carrier compositions, including various adjuvants, such as cholera-derived molecules including cholera toxin.

The targeting of liposomes can be classified based on anatomical and mechanistic factors. Anatomical classification is based on the level of selectivity, for example, organ-specific, cell-specific, and organelle-specific. Mechanistic targeting can be distinguished based upon whether it is passive or active. Passive targeting utilizes the natural tendency of liposomes to distribute to cells of the reticulo-endothelial system (RES) in organs that contain sinusoidal capillaries. Active targeting, on the other hand, involves alteration of the liposome by coupling the liposome to a specific ligand such as a monoclonal antibody, sugar, glycolipid, or protein, or by changing the composition or size of the liposome in order to achieve targeting to organs and cell types other than the naturally occurring sites of localization.

The surface of the targeted delivery system may be modified in a variety of ways. In the case of a liposomal targeted delivery system, lipid groups can be incorporated into the lipid bilayer of the liposome in order to maintain the targeting ligand in stable association with the liposomal bilayer. Various linking groups can be used for joining the lipid chains to the targeting ligand.

There are a variety of suitable formulations for the solution or suspension well known to those of skill in the art, depending on the intended use thereof.

Delivery of the described immunogenic compositions in liquid form via oral dosage has the aim of exposing the mucosa of the gastrointestinal and urogenital tracts to the immunogenic compositions. A suitable dose, stabilized to resist the pH extremes of the stomach, would deliver the immunogenic compositions to all parts of the gastrointestinal tract, especially the upper portions thereof. All means of stabilizing the immunogenic compositions in a liquid oral dosage such that the effective delivery of the composition is distributed along the gastrointestinal tract are contemplated for use with the immunogenic compositions described herein.

Delivery of the described immunogenic compositions in liquid form via ophthalmic drops has the aim of exposing the mucosa of the eyes and associated tissues to the immunogenic compositions. A typical liquid carrier for eye drops is buffered and contains other compounds well known to those of skill in the art.

Delivery of the described immunogenic compositions in liquid form via nasal drops has the aim of exposing the mucosa of the nose and sinuses and associated tissues to the immunogenic compositions. Liquid carriers for nasal drops are typically various forms of buffered saline.

Administration of the compounds discussed above can be practiced *in vitro* or *in vivo*. When practiced *in vitro*, any sterile, non-toxic route of administration may be used. When practiced *in vivo*, administration of the compounds discussed above may be achieved advantageously by subcutaneous, intravenous, intramuscular, intraocular, oral, transmucosal, or transdermal routes, for example by injection or by means of a controlled release mechanism. Examples of controlled release mechanisms include polymers, gels, microspheres, liposomes, tablets, capsules, suppositories, pumps, syringes, ocular inserts, transdermal formulations, lotions, creams, transnasal sprays, hydrophilic gums, microcapsules, inhalants, and colloidal drug delivery systems.

The compositions described herein are administered in a pharmaceutically acceptable form and in substantially non-toxic quantities. A variety of forms of the compounds administered are contemplated. The compounds may be administered in water with or without a surfactant such as hydroxypropyl cellulose. Dispersions are also contemplated, such as those utilizing glycerol, liquid polyethylene glycols, and oils. Antimicrobial compounds may also be added to the preparations. Injectable preparations

may include sterile aqueous solutions or dispersions and powders, which may be diluted or suspended in a sterile environment prior to use. Carriers such as solvents or dispersion media contain water, ethanol polyols, vegetable oils and the like may also be added to the compounds described herein. Coatings such as lecithins and surfactants may be used to maintain the proper fluidity of the composition. Isotonic agents such as sugars or sodium chloride may be added, as well as products intended to delay absorption of the active compounds such as aluminum monostearate and gelatin. Sterile injectable solutions are prepared according to methods well known to those of skill in the art and can be filtered prior to storage and/or use. Sterile powders may be vacuum or freeze dried from a solution or suspension. Sustained-release preparations and formulations are also contemplated. Any material used in the composition described herein should be pharmaceutically acceptable and substantially non-toxic in the amounts employed.

Although in some of the experiments that follow the compounds are used at a single concentration, it should be understood that in the clinical setting, the compounds may be administered in multiple doses over prolonged periods of time. Typically, the compounds may be administered for periods up to about one week, and even for extended periods longer than one month or one year. In some instances, administration of the compounds may be discontinued and then resumed at a later time.

All compound preparations may be provided in dosage unit forms for uniform dosage and ease of administration. Each dosage unit form contains a predetermined quantity of active ingredient calculated to produce a desired effect in association with an amount of pharmaceutically acceptable carrier. Such a dosage would therefore define an effective amount of a particular compound.

The immunogenic compositions described herein can be administered in amounts appropriate to those individual compounds to produce an immune response. Appropriate doses for each can readily be determined by techniques well known to those of ordinary skill in the art. Such a determination will be based, in part, on the tolerability and efficacy of a particular dose using techniques similar to those used to determine proper chemotherapeutic doses.

Additionally, a kit comprising the necessary components of a immunogenic composition that elicit an immune response to a selected immunogenic component are also contemplated.

EXAMPLE 1

Isolation and Characterization of the *csa* operon

The *csa* operon that encodes the synthesis of ETEC-CS4 pili was isolated from strain E11881A, cloned and sequenced. The *csa* operon consist of 5 contiguous genes encoding CsaA-CsaE proteins, which share homology to other pili assembly proteins, and especially to CFA/I. The *csa* operon was expressed in an attenuated *Shigella* strain, CVD1204 *guaBA*, constructing the *Shigella* expressing CS4 fimbriae vaccine strain CVD1204 (pGA2-CS4). Immunization of guinea-pigs with CVD1204 (pGA2-CS4) elicited the production of anti-CS4 antibodies that reacted with CS4 producing strains and prevented biological activities mediated by ETEC strains. This work contributes to previously reported results regarding the on expression of CFA/I, CS2, CS3 and LT in attenuated *Shigella*, emphasizing the feasibility of constructing an efficient multivalent *Shigella* – based oral ETEC vaccine.

The genes that encode the synthesis of ETEC CS4 fimbriae, *csaA,B,C,D,E,D'*, (the *csa* operon), were isolated from strain E11881A. The *csa* operon encodes a 17kDa major fimbrial subunit (CsaB), a 40 kDa tip associated protein (CsaE), a 27 kDa chaperon like protein (CsaA), a 97 kDa usher protein (CsaC), and for a deleted regulatory protein (CsaD') containing 100 amino acids out of 265. The *csa* operon is flanked by IS1E that inserted into the *csaD'* gene, and IS21 sequences upstream to the *csaA*. The *csa* operon is located on the large virulence plasmid that carries the LT genes.

A BLAST search of the predicted amino acid sequences indicated high homology of the CS4 proteins to structural and assembly proteins of CFA/I in particular, and to CS1 and CS2 fimbriae proteins. The *csaA,B,C,E* operon was cloned on a 15 copy number stabilized plasmid down stream from an osmotically regulated *omp*C promoter. Plasmid pGA2-CS4 directs in both DH5α and *Shigella flexneri* 2a strains the production of CS4 fimbriae, as detected by western blot analysis and

bacterial agglutination using anti-CS4 fimbriae immune sera. Electron microscopic examination of *Shigella* expressing the CS4 fimbriae indicated the production of lots of rod like shape extensions. Immunization of guinea pigs with *S. flexneri* 2a CVD1204 (pGA2-CS4) elicited the production of anti-CS4 antibodies that bind to CS4 fimbrial proteins, agglutinates CS4 producing strains, inhibits hemagglutination by ETEC strains and prevented the adhesion of ETEC strains to human mucosal cell line Caco-2.

Isolation of the *csa* operon

Genomic DNA of strain E11881A was isolated using GNOME DNA KIT (BIO 101, Carlsbad, CA) protocol. The DNA was partially digested with *Sau*3AI, the DNA fragments with size of 5- >20 kilobase pairs were isolated from agarose gels and ligated into vector pKS that was digested with *Bam*HI and treated with shrimp alkaline phosphatase. The ligation mix was transformed into DH5α. DH5α transformants were grown in LB in the presence of carbenicillin at 50 μg/ml or kanamycin at 50 μg/ml.

The resulting transformants were harvested into 96 well microtiter plates, and assayed for the *csa* operon by PCR tests, based on the published sequence of the *csfA* gene, (the structural gene of the CS4 fimbriae, NCBI Accession number X97493). Based on that sequence, two primers were constructed which amplified 319 base pair fragment. The primers are: CS44 bp 1 to 29: GTTGACCCTACAATTGATATTTTGCAAGC (SEQ ID NO:11) CS45 bp 378 to 348 : CGACCCCACTATAATTCCCGCCGTTGGTGC (SEQ ID NO:12). Pools of 1200 colonies were analyzed and 2 colonies were positive in the PCR test. Both colonies were found to contain the same plasmid, pKS-CSA-I, (FIGURE 1A). Sequencing of the cloned DNA fragment, indicated that plasmid pKS-CSA-I contains most of the *csa* operon, missing 430 bp from the 5' end of the *csa*A gene. DNA sequencing of the *csa* operon revealed a high degree of homology to the DNA sequences of the CFA/I operon (Table 2).

Based on that homology, two new primers were designed for screening the genomic library for the 5' end by PCR. Primer CS433 is based on the *csa*A sequences at bp 718, and primer CS434 is based on the CFA/I sequences from bp 878. CS433: GTGATATGTTTTGTTCACTTGGTAAAGATC (SEQ ID NO:13) CS434:

CTCATGGCTCCATTTGTTGCAAATGCAAACTTTATG (SEQ ID NO:14). PCR assays using these primers amplified a 429 bp DNA fragment from the genomic DNA of strain E118811A. By screening the DNA library, a positive clone that contained the entire *csa*A gene together with upstream DNA sequences was isolated. The clone contains the 8000 base pair plasmid pKS-CSA-II, (FIGURE 1B).

DNA sequence analysis

DNA sequences of the CS4 encoding genes in plasmids pKS-CSA-I and pKS-CSA-II were determined in both strands. The sequencing primers were synthesized on Perkin Elmer DNA Synthesizer model 3948 at the 40nM synthesis scale. The sequencing was performed using a Perkin Elmer DNA Synthesizer 373, with the Dye Terminators from the BIGDYE KIT, Perkin Elmer (Boston, MA). The sequencing results identified the *csa* operon and its flanking genes, as is schematically described in FIGURE 1C.

Construction of pGA2

The vector used for expression of the CS4 encoding genes in *Shigella*, pGA2, was derived from plasmid pEXO1 by replacing the *gfp* encoding gene (751 bp DNA, flanked by the restriction enzymes *Cla*I and *Nhe*I), with an 87 bp linker that contains multiple cloning sites (mcs), that was constructed by PCR, as indicated. pEXO1 is an expression plasmid derived from pGEN222. (Galen, et al., *Infect. Immun.*, 67:6424-6433 (1999)). Plasmid pGEN222 carries a two-component plasmid maintenance system comprised of the *hok-sok* post-segregational killing system plus the *parA* plasmid partitioning system. these two components have been shown to work in concert to minimize plasmid loss from a population of actively growing bacteria and to lyse any bacteria from which plasmids have segregated. pEXO1 was created from pGEN222 by replacing the *bla* gene, encoding β-lactamase, with an engineered *aph* allele encoding reduced levels of resistance to kanamycin (Blomfield, et al., *Mol. Microbiol.*, 5:1447-1457 (1991)). It was observed that expression of the wildtype *aph* allele within attenuated *S. typhi* vaccine strains lead to plasmid instability. Therefore, transcription of this *aph* allele was modified by PCR such that the separation of the -35 and -10

regions within the promoter was increased from 18 to 19 base pairs; this re-engineered allele was then introduced as an *Nhe* I fragment into pGEN222 cleaved with *Xba* I and *Spe* I to replace the *bla* gene.

The PCR fragment encoding for *mcs* was used as a *Cla* I - *Nhe* I restriction fragment to replace the *Cla* I-*Nhe* I restriction fragment within pEXO1 encoding the *gfpuv* allele, creating pGA2. pGA2 is therefore expected to be present at approximately 15 copies per chromosomal equivalent, and to drive expression of CS4 encoding genes from the osmotically responsive *ompC* promoter.

The *mcs* was constructed by PCR using the following primers: Primer 4a: GGGATCGATCCCGGGGCGGCCGCGGGCCCGGTACCAGGCCTTCTAGAAAGC TTGACGTCG (SEQ ID NO: 15); primer 4b: CCCGCTAGCGGCGCGCCTCGCGAGGATCCGTCGACGACGTCAAGCTTTCT AGA AGGCCTGG (SEQ ID NO: 16); Primer 4c: AAGCTTGACGTCGTCGACGG (SEQ ID NO: 17); Primer 4d: CCCGCTAGC GGCGCGCCTCGCG (SEQ ID NO: 18).

Construction of the CS4 encoding plasmid

Plasmid pGA2-CS4 (FIGURE 2B), was constructed in three steps. The 5' end of the *csa*A gene was cloned from plasmid pKS-CSA-II on a 710 bp *Hpa*I / *Xba*I fragment, into pKS *Sma*I / *Xba*I. The *Hpa*I site is located 241 bp upstream the ATG codon for the *csa*A gene. The DNA fragment was then further cloned as a *Kpn*I / *Xba*I fragment into pGA2 *Kpn*I / *Xba*I, to construct pGA2-Δ*csa*A plasmid. The remaining CS4 encoding genes, (the 3' end of *csa*A, *csa*B, *csa*C, and *csa*E) were cloned from pKS-CSA-I as a 4526 bp *Xba*I fragment. The two *Xba*I sites are located in the *csa*A gene, and in the stop codon of the *csa*E gene. The DNA fragment was ligated to pGA2-Δ*csa*A *Xba*I site, to construct plasmid pGA2-CS4.

Additional DNA primers for PCR assays

For location of different genes by PCR assays, the following primers were constructed. For LT gene (AB011677, GI 3062900): LTA bp 162:CCGTGCTGACTCTACACCCCCAGATG (SEQ ID NO: 19) and LTB bp 895:GCACATAGAGAGGATAGTAACGCCG (SEQ ID NO:20). For *gyrA* gene

(X57174, GI 41641): *GYRA* bp 347:CGGTCATTGTTGGCCGTGCGCTGCC (SEQ ID NO:21) and *GYRA* bp 1147:CACGCAGCGCGCTGATGCCTTCCACGCG (SEQ ID NO:22). For *csaD* gene: CS4D3 : from *csa* operon bp 5540-5564, and *cfaD* (GI 145505) bp 1236-1260. CATATTTGATATCTGAGATATCTGG (SEQ ID NO:23) CS4D2: From *csa* operon bp 6005- 60030, and *cfaD* bp 771-794, TGTTGCATTCAGATTGAACGGAG (SEQ ID NO:24). CS4D1: From *cfaD* bp 606-629, at similar region to *rns*. This DNA region is missing in *csa* operon : TATTATGATTCATAAATACACTGT (SEQ ID NO:25). PCR assays using primers CS4D2/CS4D3 is expected to amplify a 476 bp DNA, and with primers CS4D1/CS4D3, a 646 bp DNA.

Transformation of *Shigella* strains

Shigella strains were grown in trypticase soy agar (TSA) agar plates containing 0.1% Congo red and 10 μg/ml guanine. Competent cells of *S. flexneri* 2a CVD1204 were prepared by growing the cells in L broth supplemented with guanine to OD_{600} nm of 0.6. The cells were precipitated, washed twice with cold H_2O, once with cold 10% glycerol and resuspended in the same buffer at 1/100 of the original volume. A mixture of 150 μl cells and plasmid DNA were electroporated in a 0.2 cm curettes in a Gene Pulser (BioRad Laboratories, Hercules, CA) using 2.5 kV, 200Ω, 25 μF, or 1.75kV, 600Ω, 25 μF. Transformants were selected on kanamycin, guanine and Congo red containing TSA plates.

Detection of pili synthesis

ETEC strains were grown on CFA (Hamers, et al., *Microb. Pathog.*, 6:297-309 (1989)) plates at 37°C, and the bacteria was resuspended in phosphate buffered saline (PBS). *Shigella* strains that contain the plasmid pGA2-CS4 were grown in TS broth, [Trypton (Difco), 1.5%; Soyton (Difco), 0.5%)], containing 0, 50, 150 and 300 mM NaCl to a logarithmic phase growth. The bacteria were assayed for pili production by bacterial agglutination assays, and by immunoblotting of cell extracts. For immunoblotting the bacterial cultures were adjusted to OD_{600nm}=10 and boiled for 10 minutes in Laemnli sample buffer (BioRad). The cell extracts proteins were separated

on SDS-PAGE (15%), transferred to nitrocellulose filters (MSI) and probed with anti-CS4 serum. The specific anti-CS4 antiserum was produced in rabbits by immunization with ETEC strain E11881A (a $CS4^+$ $CS6^+$ producer strain), and absorption of the sera on EII88IC (a $CS4^-CS6^+$ strain).

Hemagglutination

For hemagglutination tests (Willshaw, et al., *FEMS Microbiol. Lett.*, 49:473-478 (1988)), ETEC and *Shigella* strains were grown overnight on CFA or TSA/ CR/guanine / kanamycin containing plates, respectively, and were resuspended in PBS to $OD_{600\ nm}$ of 10. The slide hemagglutination tests, were performed as described by Sakellaris, et al 1999 (Sakellaris, et al., *Proc. Natl. Acad. Sci. U.S.A.*, 96:12828-12832 (1999)), by mixing 20 µl bacterial suspension with 20µl PBS containing 0.1M D-mannose and 20 µl washed human erythrocytes of group A. For hemagglutination inhibition assays, the bacterial suspensions were incubated with four-fold diluted antibodies for 1 hour at 37°C prior to the hemagglutination tests.

Adhesion and inhibition of adhesion to Caco-2 cells

Caco-2 cells were grown for 15 days in 8 chamber slides in Dulbecco's modified Eagels medium containing fetal calf serum (20%) and glutamine (1%). Adhesion assays were performed by incubating washed cells with bacterial concentrations of 10^7 and 10^8/ml, for 2 and 3 hours in DMEM containing 0.1M D-mannose, (Viboud, et al., *Microb. Pathog.*, 21:139-147 (1996)). In the inhibition of adhesion assays, the bacterial cultures were incubated with antibodies for 1 hour, at 37°C prior to their addition to the washed Caco-2 cells. Following the incubation, the Caco-2 cells were washed 5 times with PBS, fixed with methanol and stained with Giemsa (10%).

Immunization

Guinea pigs, anesthetized subcutaneously with ketamine HCL (40mg/kg) and xylazine (5mg/kg), were inoculated intranasal administration twice on days 0 and 14, with ~$2x10^9$ bacteria (0.1 ml of $40OD_{600nm}$) that grew on TSA / Congo red / guanine containing plates and harvested in PBS. Two groups of animals were immunized:

group I with CVD1204 (pGA2), and group II with CVD1204 (pGA2-CS4), a CS4 producing strain. Sera were obtained on days 0, 14 and 30 by anterior vena cava puncture of anesthetized animals.

Results

Cloning and sequencing of the CS4 fimbriae encoding genes

The *csa* operon that encode the synthesis of the CS4 fimbriae, was isolated from a genomic DNA library of strain E11881A, as described above. Most of the *csa* operon was cloned on plasmid pKS- CSA-I that contains the carboxy terminal part of the *csaA* gene, the *csaB, csaC, csaE* genes and a disrupted *csaD'* gene. The *csaD'* was disrupted by integration of an IS1 element, creating a deletion of the amino terminal 48 amino acids, and a frame shift mutation that resulted in stop codon. Agglutinations assays of DH5α (pKS-CSA-I) strains with rabbit serum anti E11881A (that was absorbed on strain E11881C) were positive, which indicated expression of the CS4 fimbriae. Western blot tests indicated the presence of two fimbrial bands, the mature 17 kDa protein, and a higher MW protein, probably the pre cleaved form. By further screening of the CS4 library plasmid pKS-CSA-II was isolated. The plasmid contains the *csa* promoter site, the *csaA* gene and the amino terminal region of *csaB*. Up stream of the *csaA* gene is an IS 21 element. The *csa* operon is located on a ~10,500 bp DNA fragment that is flanked by insertion elements, similar to a pathogenicity island, as schematically described in FIGURE 1C.

7,239 bp of the CS4 pathogenicity island was sequenced in both directions, indicating a 34.88% G+C region. The *csa* operon is located between bp 1 and 6,095 bp. It contains 5 ORF, 4 genes, *csaA, csaB, csaC*, and *csaE* are transcribed in the same direction down stream from a predicted promoter site, and *csaD'* from the antiparallel strand. The location of each gene is described in Table 1, and schematically presented in FIGURE 1C. A BLAST search for homology of the *csa* DNA sequence (bases 1 to 6096) to other genes indicated homology to CFA/I, CS1 and CS2 ETEC fimbriae encoding genes, to the structural genes of CS4 and CS14, and to fimbriae regulatory genes *cfa*D, *rns*, *csv*R and *agg*R, as presented in Table 2. The results of the BLAST search indicate that the *csa* operon has a high DNA similarity to *cfaI* operon; 5,420 bp

out of 6,069 bp are >91% similar. Comparing the DNA sequence of each gene to other fimbriae encoding genes indicate the following results:

csaA: 93% of the 717 bp of the *csaA* DNA sequences are identical to 716 bp of *cfaA* (M55661.1), and 44 bp from its 5' end are 90% identical to *csoA* and *cooB* 713 bp genes (X62879.1 and X62495.1).

csaB: 402 bp of *csfA* (X97493.1), the published DNA sequence of the CS4 structural gene, are 99% identical to 501 bp of the *csaB* DNA sequences. Of the 512 bp of *cfaB*, (M55661.1), the sequence of 107 bp is 93% identical to *csaB*. 93 bases from the 5' end of *csuA1* 504 bp, (X9749.1), the structural gene for CS14 fimbriae, are 89% identical to *csaB* . 35 bp from the 5' end of *cotA* 512 bp gene (Z47800.1) is 94% identical to *csaB*.

csaC : 2601 bp of *csaC* gene share 96% DNA sequences similarity to 2507 bp of *cfaC* (M55661.1) , 80% homology to 391 bp out of 2618 bp of *coo*C (X76908.1), and 85% homology to 124 bp out of 2597 bp of *cotC* (Z47800.1) genes.

csaE : 1086 bp of the *csaE* DNA sequences are ≥ 84% identical to 999 bp out of 1082 bp of *cfaE* (M55661.1), and 32 bp out of 1091 bp of *coo*D (X76908.1) are 90% identical to *csaE* .

csaD' : *csaD'* DNA sequence is similar to *cfa*D' (CFA/I), *cfa*D, *rns* and *csv*R from ETEC strains, and *agg*R from EHEC fimbriae. DNA sequence similarity between *csaD'*, *cfa*D and rns is of 751/817 bp (91%) for *cfa*D, and 721/787 (91%) for *rns*.

Comparing the regions of similarity in both genes indicated that *csaD'* gene is missing 143 bp from its 5'- end because of an insertion of an IS element. Blast search for DNA sequences downstream the *csa* operon indicated that from base 6096 to 6870 there is homology to IS1E sequences, (NCBI X52537, identities = 765/769, 99%), from bp 6892 to 7054, there is homology to EPEC strain plasmid pB171 (GI 6009376), between bp 16240 and 16471. And from bp 7062-7239 there is homology to *Shigella sonnei* strain P9 plasmid colIb (GI 4512437) between bp 4895 and 5568.

Promoter site

The promoter site for the *csa* operon was predicted using the "Promoter Prediction By Neural Network", and it is proposed to be located between bp 145 to 194, 89 bp upstream the ATG codon for *csa*A gene. The predicted promoter sequences is as follow:

TGTGGGTATTTGTTTGGACATCGCAGCATTAAATATAAAAATAGCACAGG (SEQ ID NO:26). The large underline "A" is the predicted transcription start, and the shaded nucleotides are the possible -10 and -35 sequences. A BLAST search indicated a 28/31 bp (90%) homology to the CFA/I fimbrial operon, at between bp 687 to 717, 131 bp upstream to the *cfaA* gene.

CS4 pili structural and assemble proteins. The *csa* operon encodes the synthesis of five proteins: CsaA, CsaB, CsaC, CsaE and CsaD. The location of the genes and the size of the putative proteins are described in Table 1. According to sequence similarity of the CsaA-E proteins to other ETEC fimbriae proteins, the predicted functions of the CsaA-E proteins is as follow:

CsaA function as a periplasmic chaperon like protein. PSORT analysis indicate that the putative protein is a bacterial periplasmic space protein (certainty of 0.939). CsaB is the major pilin subunit. CsaC is a membrane usher protein. A PSORT analysis predicted an outer membrane location (certainty of 0.926). CsaE is assumed to be at the fimbriae tip. A PHDsec analysis (for the prediction of secondary structure) indicated that the protein is a compact protein as a globular domain. CsaD' is a fimbriae regulatory protein. The CsaD' protein contains 100 amino acids from the carboxy terminal part of the protein, missing the first 48 amino acids (based on homology to *cfa*D gene from CFA/I pili), because of an insertion of an IS1E element. Following the 100 amino acid is a frame shift mutation that encodes for a stop codon.

Homology of the CsaA-E proteins to other fimbriae proteins

A BLAST search with the putative amino acid sequence of CsaA-E proteins indicated homology to fimbriae proteins from ETEC strains and *Salmonella typhi* as described in Table 3. The amino acid sequence of the structural and assemble proteins of the CS4 fimbria have similarity to ETEC proteins producing the CFA/I, CS1 and

CS2 fimbriae, and to the *Salmonella* fimbria. In addition, the CS4 structural proteins have amino acid sequences similar to the structural proteins of CS14, CS17, CS19 and *B. cepacia*. The CS4 fimbriae structural protein CsaB and the tip-associated protein CsaE are responsible for the fimbrial structure and for the bacterial attachment to intestine cells. An alignment of the amino acids sequences of the fimbrial proteins that are similar to CsaB and CsaE is described in FIGURES 3 and 4.

TABLE 3

Similarity in AA sequences between ETEC CsaA-E proteins and other fimbrial proteins

CS4 protein	Pili	Protein	#amino acids	Identities		region in CS4 proteins	region in compared protein	Gene no.
CsaA	CFA/I	CfaA	238	208/238	(87%)	1-238	1-238	gi 145508
	CS1	CooB	238	124/221	(56%)	1-219	1-219	gi 95719
	CS2	CotB	238	111/238	(46%)	2-238	2-238	gi 897726
		TsaA[1]	236	60/194	(30%)	17-207	22-204	gi 5640159
	CS5	25.9KD	224	52/218	(23%)			P33792
CsaB	CS4	CsfA	134	109/134	(81%)	34-160	1-134	gi 1304302
	CFA/I	CfaB	170	99/170	(58%)	1-167	1-170	GI 145509
	CS14	CsuA1	168	95/164	(57%)	1-164	1-164	gi 1304304
	CS1	CooA	171	85/168	(50%)	1-164	1-168	gi 78442
	CS2	CotA	170	76/167	(45%)	1-164	1-167	gi 897727
	CS14	CsuA2	142	67/132	(50%)	34-164	1-132	gi 1304306
	CS19	CsdA	133	56/131	(42%)	34-164	1-130	gi 1304300
	CS17	CsbA	135	54/132	(40%)	34-164	1-132	gi 1304298
		TsaB[1]	191	58/169	(34%)	1-166	22-187	gi 5640160
	Cable	CblA[2]	184	50/150	(33%)	20-167	13-159	gi 606843
CsaC	CFA/I	CfaC	869	800/868	(92%)	16-883	2-869	gi 2121079
	CS1	CooC	872	531/841	(63%)	40-880	31-868	gi 488736
	CS2	CotC	866	469/839	(55%)	40-878	29-864	gi 897728
	CS6	CssD		120/549	(21%)			P53513
		TsaC[1]	895	260/867	(29%)	42-872	25-882	gi 5381204
		AtfC[3]	843	111/491	(22%)	211-670	207-665	gi 1504107
		CaflA[4]	833	100/511	(19%)	209-685	205-661	gi 3883097
CsaE	CFA/I	CfaE	360	268/361	(74%)	2-362	1-360	gi 2121080
	CS1	CooD	363	177/329	(59%)	41-362	39-363	gi 488737
	CS2	CotD	364	161/339	(47%)	29-362	28-364	gi 897729
		TsaD[1]	359	90/303	(29%)	71-361	79-358	gi 5640162
		rfbF[5]	187	19/66	(28%)	146-211	55-114	gi 48590

TABLE 3
(Continued)

CS4 protein	Pili	Protein	#amino acids	Identities		region in CS4 proteins	region in compared protein	Gene no.
CsaD	CFA/I	CfaD	265	90/101	(89%)	9-109	49-149	gi 145506
	CFA/I	CfaD'	144	90/101	(89%)	9-104	49-144	gi 145508
		RNS	265	89/101	(88%)	9-104	49-149	gi 145512
		CSVR	305	75/103	(72%)	9-104	49-151	gi 95726
	AAF/I	AGGR	265	57/101	(56%)	9-109	49-149	gi 420983

[1] *Salmonella typhi*
[2] *Burkholderia cepacia*
[3] *Proteus mirabilis* outer membrane usher protein
[4] *Yersinia pestis* F1 capsule anchoring protein
[5] *Yersinia enterocolitica*

Localization of the csa operon in strain E11881A

In order to identify the location of the *csa* operon, total genomic DNA was isolated from strain E11881A, and was subjected to agarose gel electrophoresis. The electrophoresis results indicated the presence of 3 plasmids: a large plasmid located above the chromosomal band, and two smaller plasmids located under the chromosomal DNA band. The plasmids and the chromosomal bands were gel eluted and tested by PCR assays for the presence of the *csa* operon, using primers CS44/CS45 for amplification of the *csaB* gene. These results indicated that the *csa* operon was located on the large plasmid. The location of the LT gene was detected by PCR assays using primers LTA162/LTB895 (for amplification of a 708 bp DNA), indicating the LT encoding genes were located on the large plasmid. Amplification of the *gyr*A gene, using primers GYRA347/GYRA1147 (for amplification of a 748bp DNA) indicated a chromosomal location. In order to see whether strain E11881A contained a complete *csaD* gene, PCR assays were performed with primers that are homologous to *csaD'* gene (CS4D2 and CS4D3), and with primer CS4D1 that is homologous to *cfaD* and *rns* genes (CSD41/CSD43). The results indicates that strain E11881A does not contain the complete cfaD gene. Primers CS4D2/CS4D3 amplified the expected 476 DNA fragment, while primers CS4D1/CS4D3, did not amplified an expected 646 bp. Strains EII881E, DS9-1 H10407, and C91f, contained the complete gene.

Expression of the CS4 pili in DH5α and CVD1204

The expression of the CS4 fimbriae was detected following cloning of 5281 bp of the *csa* operon in vector pGA2, to construct the plasmid pGA2- CS4, (FIGURE 2B). The cloned *csa* operon contained 240 bp of the promoter region up stream the ATG codon for *csaA* gene, the *csaA, csaB, csaC* and *csaE* genes. The *csa* operon was cloned downstream of the *omp*C promoter, which is osmotically regulated. Plasmid pGA2-CS4 was transferred to *E.coli* DH5α and *Shigella* CVD1204 strains. CS4 fimbriae production by strains DH5α (pGA2-CS4) and CVD1204 (pGA2-CS4) strains was detected by agglutination assays using rabbit serum against the CS4 fimbriae and western blot of cell extracts. Western blot results indicates that the cloned *csaABCE* gene cluster encodes for a 17 kDa band that correspond to the CS4 fimbriae.

Antibody response of guinea-pigs to CVD1204 (pGA2-CS4)

To test whether the cloned *csa* operon could induce an immune response against ETEC-CS4, guinea pigs were immunized by two intranasal administrations of $2x10^9$ live CVD1204 (pGA2-CS4), and as control with CVD1204 (pGA2) strain. The immunized animals were tested for bacterial agglutination, immunoblotting, and inhibition of hemagglutination and adherence of CS4 expressing strains to Caco-2 cells.

Agglutination assays

Bacterial agglutination assays done with immune guinea pigs indicated that immunized animals with CVD1204 (pGA2-CS4) developed antibodies that agglutinates all CS4 producing strains, and especially the high fimbriae producing strains DS 9-1, DH5α(pGA2-CS4) and CVD1204(pGA2-CS4), as described in Table 4. The end point of serum dilution that agglutinates DS 9-1 and DH5α(pGA2-CS4) strains is >1:100 and >1:1000, respectively. The control sera from animals immunized with CVD1204 (pGA2) strain, agglutinate CVD1204 strains. Bacterial agglutination assays using rabbit anti-CS4 antiserum were also performed and the data from these assays is presented in Table 5.

TABLE 4
Bacterial agglutination assays by rabbit anti CS4 antiserum

Strain	NaCl mM	End point dilution of the antibody that result in bacterial agglutination	Induction factor
CVD1204(pGA2-CS4)	0	1:400	
	50	1:800	2
	150	1:1600	4
	300	1:800	2
CVD1204(pGA2)	0	<1:10	
	150	<1:10	none

TABLE 5
Bacterial agglutination assays by immune guinea pigs serum

Tested strain[1]	Relevant phenotype	Intensity of bacterial agglutination by immune guinea pigs sera[2] against:	
		CVD1204 (pGA2-CS4)	CVD1204 (pGA2)
E11881A	$CS4^+$ $CS6^+$	+	-
E11881E	$CS4^+$ $CS6^+$	+	-
EII881C	$CS4^-$ $CS6^+$	-	-
E11881/G28	$CS4^-$ $CS6^-$	-	-
DS9-1	$CS4^+$ $CS6^+$	++++	-
DH5α(pGA2-CS4)	$CS4^+$	++++	-
DH5α	$CS4^-$	-	-
CVD1204(pGA2-CS4)	$CS4^+$	++++	+
CVD1204	$CS4^-$	+	++

[1] ETEC and *Shigella* strains were grown on CFA and TSA, guanine, Km plates, respectively; and resuspended to OD_{600}nm of 5.

[2] The assays were performed with antibodies dilution of 1:10 and higher.

Immunoblotting assays

The immunized guinea pigs were tested in immunoblotting experiments to bind to CS4 fimbriae proteins. The assays were performed by probing CS4 producing strains with the guinea pig immune serum and rabbit anti-CS4 serum. Cell extract were isolated from ETEC strains E11881A and DS 9-1, that grew on CFA plates at 37°C and 22°C (a condition that suppresses fimbriae production), from DH5α (pGA2-CS4) and CVD1204 (pGA2-CS4) strains that grew in LB broth in the presence of Km. The immunoblotting results showed that guinea pigs serum from CS4 immunized animals

reacted with a 17 kDa band protein which is the CS4 fimbrial structural protein. ETEC strains produced a 17 kDa band at 37°C but not at 22°C. *E.coli* and *Shigella* strains that contains the plasmid pGA2-CS4 produced 17 kDa protein bands, which was not produced in the corresponding untransformed cells.

Inhibition of hemagglutination

ETEC strains that produce the CS4 fimbriae cause a mannose resistant agglutination of human red blood cells type A. The ability of ETEC-CS4 strains, DH5α, and a CVD1204 strain that contains the cloned CS4 encoding genes, were tested for hemagglutination. The results are presented in Table 6. The hemagglutination assays indicated that all CS4 pili producing strains caused hemagglutination. ETEC-CS4 strains demonstrate a more intense hemagglutination than strains expressing the cloned *csa* operon. Non CS4 producing strains did not caused any hemagglutination.

TABLE 6
Efficiency of bacterial induced hemagglutination

Tested strains	Relevant phenotype	Hemagglutination
E11881A	$CS4^+$ $CS6^+$	++
E11881E	$CS4^+$ $CS6^+$	+++
EII881C	$CS4^-$ $CS6^+$	-
DS 9-1 (37°C*)	$CS4^+$ $CS6^+$	++++
DS 9-1 (22°C*)	$CS4^-$	-
DH5α(pGA2-CS4)	$CS4^+$	+
DH5α	$CS4^-$	-
CVD1204(pGA2-CS4)	$CS4^+$	+
CVD1204(pGA2)	$CS4^-$	-

* Temperature of growth.

The guinea pigs anti-CS4 antiserum was tested for inhibition of hemagglutination. ETEC-CS4 strains were incubated with various antibody dilutions prior to their addition to the red blood cells suspension. The results that are presented in Table 7 indicates that the guinea pigs anti-CS4 antibodies inhibits the hemagglutination, while the control immune serum against the *Shigella* has no effect on the ETEC-CS4 induced hemagglutination.

TABLE 7
Efficiency of inhibition of bacterial induced hemagglutination by guinea pigs anti-CS4 antibodies

Tested strains	End point dilution of antibodies that inhibit bacterial induced hemagglutination. Antibodies were from guinea pigs immunized with:	
E11881A	1:256	<1:2
DS 9-1	1:32	<1:2

Inhibition of adherence of ETEC-CS4 strains to Caco 2 cells

ETEC-CS4 producing strains have the ability to adhere to human carcinoma cell line Caco-2. The adhesion of ETEC strain DS 9-1to the Caco-2 cells, and the effect of the guinea-pigs anti-CS4 antibodies on this adhesion were tested. The results indicate that ~10-100 bacteria adhere to the Caco-2 cells. As a control, the adhesion of DH5α bacteria was tested, and only very few bacteria were found to adhere. Preincubation of the ETEC DS 9-1 with guinea pigs anti-CS4 serum (1:10 dilutions), totally inhibit the adherence of the bacterial cells to the Caco-2 cells.

EXAMPLE 2

Construction of an Attenuated *S. typhi* Strain that Constitutively Expresses CS4

In order to change the expression of CS4 in pGA2-CS4 from osmotically regulated to constitutive, a strong promoter, *e.g.*, P_{tac}, is used. The promoter P_{tac} is constitutively active in Salmonella spp., as these organisms lack laqI. Accordingly, constitutive CS4-expressor derivatives of CVD 915 are constructed in the manner described below.

First, the *omp*C promoter is removed from the pGA2-CS4 construct and replaced with the P_{tac} promoter, resulting in the creation of pGA2tac-CS4. After confirming the presence of the P_{tac} promoter in pGA2tac-CS4, the construct is introduced into *S. typhi* strain CVD-915, using standard molecular biology techniques.

Constitutive expression of the CS4 antigen is assessed by the agglutination assay described in Example 1. Using wild-type *S. typhi* strain Ty2 as the positive control, and

untransformed CVD-915 (CVD 915$^-$) as the negative control, expression of the CS4 antigen is found to be strong, constitutive, and not regulated by changes in osmolarity.

EXAMPLE 3

Immune Response Against Constitutively Expressed CS4

Groups of ten 6 weeks old Balb/c mice are immunized intranasally with $1.0x10^{10}$ cfu of strain CVD 915 with pGA2tac-CS4 (CVD 915$^+$) or CVD 915 without pGA2tac-CS4 (CVD 915$^-$). Mice are bled prior and 30 days after their immunization, and their serum is stored at -20°C until use. Antibodies present in the serum against *S. typhi* LPS, H (flagella) and CS4 antigens are determined by ELISA. The results indicate that immunization with strain CVD 915$^+$ elicits antibody levels against the CS4 antigen that are significantly higher than those obtained with strain CVD 915$^-$. The immune responses against other *S. typhi* antigens (LPS and H) are similar between both immunized groups. The results demonstrate that the constitutive expression of the CS4 antigen enhances the immune response against this antigen without interfering with the immune response against other somatic *S. typhi* antigens.

EXAMPLE 4

Immune Response Against Constitutively Expressed CS4

The enterotoxic *E. coli* strain expressing the CS4 antigen (ETEC-CS4) from Example 1 is formulated into an immunogenic composition for nasal administration into a human subject (ETEC-CS4). Approximately $1.0x10^{10}$ cfu of the recombinant bacteria ETEC-CS4 is nasally administered to a human subject. A booster is administered two weeks subsequent to the first administration. Blood is subsequently drawn from the subject and assayed for the presence of anti-ETEC-CS4 and anti-CS4 antibodies. Such antibodies are found in the blood sample.

Although the invention has been described with reference to embodiments and examples, it should be understood that various modifications can be made without departing from the spirit of the invention. Accordingly, the invention is limited only by the following claims. All references cited herein are hereby expressly incorporated by reference.